

**THE INFLUENCE OF PROGRAM THEORY-BASED EVALUATION ON THE USE OF  
EVALUATION INFORMATION:  
A LONGITUDINAL STUDY**

Sheri L. Hudson

BSc (Hons), *E.Mont.Coll.* PgDipSocResEval, *Murdoch*  
MEd (1<sup>st</sup> Class Hons), *Murdoch*

This thesis is presented for the degree of Doctor of Philosophy  
Murdoch University  
2008

## **DECLARATION**

I declare that this thesis is my own account of my research and contains as its main content, work which has not previously been submitted for a degree at any tertiary education institution.

---

Sheri Lee Hudson

## **ACKNOWLEDGEMENTS**

In particular, I would like to thank my principal supervisor, Dr. Ralph Straton, for his gentle guidance, patience, perseverance and understanding throughout this journey. Your encouragement, knowledge-sharing and dedicated mentoring is greatly appreciated.

My grateful thanks are extended to Dr. Richard Cummings, my associate supervisor, who has been central in moving this research along to completion. His continual assistance, support, and critical comments have been very much appreciated.

This study would not have been possible without the co-operation of those connected with the evaluation studies which were the focus of the investigation. Many people made themselves available for lengthy interviews in which they provided thoughtful responses. They were tolerant of the detailed questioning involved in the conduct of this investigation, and willingly gave of their time. Their assistance is gratefully acknowledged.

I would like to thank Mr. Steven Newman for proof-reading the entire thesis not once, but twice.

My deep appreciation is extended to my many friends who accompanied me in this journey. It would have been a very lonely road without their support and encouragement to light my way.

I am sincerely and deeply grateful for my three sons, Christian, Cole and Connor, for keeping me grounded and insisting I complete this journey.

Lastly, I will pay tribute to my Gramma, who passed away during this work, for teaching me everything she was suppose to.

## **ABSTRACT**

Theory-based evaluation approaches have been promoted by program evaluators as enhancing program-related knowledge and decision-making. Although a substantial amount of conceptual work related to theory-based evaluation studies has been undertaken in the past thirty years, little empirical research has been completed to investigate the influence of a theory-based program evaluation on utilisation of the evaluation's information. The research reported here investigated the relationship between the degree to which program theory was used as the basis for an evaluation study, and the nature and extent of the utilisation of the information resulting from the study.

A model of factors thought to influence the use of program evaluation studies and the information they yield was developed and investigated. The 'program theory' factor was concerned with the influence of the use of both 'causative' program theory and 'implementation' program theory in an evaluation on the use of the evaluation information, the primary variable considered by this study. The model also includes the main factors found to facilitate the utilisation of evaluation information, identified in earlier empirical research by Alkin and Associates (1985), Cousins and Leithwood (1986, 1993), Cummings (1997), Leviton and Hughes (1981), Hudson-Mabbs (1993), Pawson and Tilley (1997) and Preskill and Torres (1997). The other factors include the characteristics of the learning environment present in the organisation responsible for the delivery of the program, the stakeholders of the evaluation, the evaluation team responsible for undertaking the evaluation, the evaluation study and its processes, stakeholder involvement in the study, stakeholder commitment to the study (pre- and post-study), stakeholder involvement in the program theory elaboration process, and program theory use in the final report. The logic of the conceptual model underpinned and guided the design, methods, instrument development, data analysis and structural model development for this research.

Concurrent empirical case studies of three program evaluation studies were undertaken with a view to identifying the extent of theory use and the utilisation of the information from each evaluation. The study adopted a longitudinal design and used structural equation modelling to analyse the model. Qualitative data were used to gain further insight into the study findings.

Although this investigation has not been able to confirm that a greater use of information is associated with a greater use of program theory in an evaluation study, interesting interactions between program theory and other predictor variables, such as the characteristics of the evaluation study have been identified. Furthermore, this research provides insight into the vulnerability of an evaluation study to contextual factors, which are often outside the control of the evaluation team. It also provides further evidence of the importance of stakeholder involvement in an evaluation study and the extent to which stakeholders are influenced by the evaluation information.

## TABLE OF CONTENTS

Declaration	i
Acknowledgements	ii
Abstract	iii
Table of Contents	v
Table of Figures	ix
List of Tables	xi
 Chapter I	
Introduction	1
Modern Program Evaluation	2
Program Theory-Based Evaluation	3
Evaluation Utilisation	5
Significance of the Study	6
Purpose of the Study	7
The Study Method	8
Organisation of the Dissertation	9
 Chapter II	
Literature Review	12
Emergence of the Theory-Based Evaluation Approach	13
Program Theory	23
The Role of Context	33
The Use of Evaluation Information	39
Conclusion	60

Chapter III	
Model Development	62
Introduction	62
The Evaluation Process	63
Descriptive Model of Program Theory Development	75
Theory-Based Data Collection Matrix	88
The Form of the Model to be Tested	97
Summary	104
Elaboration of the Structural Model	105
Classification of Program Theory Use in Evaluation Studies	108
Measure of Program Theory Use Classifications	155
Conclusion	157
 Chapter IV	
Methods of the Investigation	159
Study Focus and design	159
Design of The Study	160
Sources of Information	169
Data Collection Instruments	176
Procedure	190
Data Analysis	192
Conclusion	193
 Chapter V	
Development and Analysis of Instruments	195
Operationalisation of the Latent Variables	196
The Organisational Environment Interview Schedule	199
Stakeholder Interview Schedule I	220
Evaluator Interview Schedule I	240
Process Use Interview Schedule	265
Stakeholder Interview Schedule II	268
Evaluator Interview Schedule II	296
Conclusion	302

Chapter VI	
The Evaluation Studies	303
Youth on Health Festival	305
The Pre-Apprenticeship Training Program in Western Australia	316
Making Consistent Judgments Professional Learning Module	325
The Classification Matrix	338
Summary	341
 Chapter VII	
Findings of the Study	342
Structural Equation Modeling	342
Model Fit	345
Testing of the Measurement Model	350
Testing of the Structural Model	365
Further Development of the Core Model	388
Conclusion	411
 Chapter VIII	
The Effect of Program Theory on Evaluation Information Use	413
The Core Model	414
Development of the Fully Elaborated Model of the Study	433
Conclusion	442
 Chapter IX	
Summary and Conclusions	445
A Synopsis of the Study	445
Conclusions of the Study	448
Strengths and Weaknesses of the Study Design	453
Implications for Future Research	456
Implications for Evaluation Practice	458
Conclusion	462



Reference List	464
Appendix A	484
Information Letter for Interviewees	
Consent Form to be signed by Interviewee	
Organisational Environment Characteristics Interview Schedule	
Appendix B	497
Information Letter for Interviewees	
Consent Form to be signed by Interviewee	
Stakeholder Interview Schedule I	
Process Use Interview Schedule	
Stakeholder Interview Schedule II	
Appendix C	538
Information Letter for Interviewees	
Consent Form to be signed by Interviewee	
Evaluator Interview Schedule I	
Evaluator Interview Schedule II	
Appendix D	566
Instrument for developing a scale to rate the degree of program stake or interest of various stakeholder roles.	

## TABLE OF FIGURES

Figure 3.1	Theory Use of Evaluations Targeted for Inclusion in Study	63
Figure 3.2	A Model of the Evaluation Process	67
Figure 3.3	Theory-Based Evaluation Process Stages	68
Figure 3.4	Theory Development for a Program: Identified Need, General Program Participants	77
Figure 3.5	Theory Development for a Program: Program Goals, General Program Participants, Causative Theory, and Implementation Theory	78
Figure 3.6	Theory Development for a Program: General Program Specification Development	79
Figure 3.7	Theory Development for a Program: Local Program Specification Development	81
Figure 3.8	Theory Development for a Program: The Program	83
Figure 3.9	Theory Development for a Program: Project Evaluation Conceptual Utilisation	84
Figure 3.10	Theory Development for a Program: Project Evaluation Instrumental Utilisation	86
Figure 3.11	Descriptive Matrix presented by Stake (1967; p. 529) as a portion of his Data Collection Matrix.	90
Figure 3.12	Core Structural Equation Model	100
Figure 3.13	Intended Version of the Elaborated Structural Model	107
Figure 4.1	Study Design Diagram	160
Figure 4.2	Study Procedure	194
Figure 5.1	Question Format	289
Figure 6.1	YOHFest Program Theory Representation	307
Figure 6.2	Pre-Apprenticeship Program Theory Representation	322
Figure 6.3	MCJ Program Theory Representation	337
Figure 7.1	Organisational Characteristics Measurement Model	357
Figure 7.2	Evaluator Characteristics Measurement Model	358
Figure 7.3	Stakeholder Characteristics Measurement Model	360
Figure 7.4	Evaluation Characteristics Measurement Model	361
Figure 7.5	Importance/ Likelihood of Use Measurement Model	364
Figure 7.6	Generic Version of Core Structural Equation Model	374
Figure 7.7	Structural Model 1a	379
Figure 7.8	Structural Model 1b	379

Figure 7.10	Structural Model 2b	383
Figure 7.11	Fully Elaborated Model	389
Figure 7.12	Core Model 1a Elaborated with OLE Variable	391
Figure 7.13	Core Model 1b Elaborated with OLE Variable	391
Figure 7.14	Core Model 2a Elaborated with OLE Variable	394
Figure 7.15	Core Model 2b Elaborated with OLE Variable	394
Figure 7.16	Model Version 1a Elaborated with Evaluator Characteristics	399
Figure 7.17	Model Version 1b Elaborated with Evaluator Characteristics	399
Figure 7.18	Model Version 2a Elaborated with Evaluator Characteristics	402
Figure 7.19	Model Version 2b Elaborated with Evaluator Characteristics	402
Figure 7.20	Core Model 1a Elaborated with Involvement Characteristics	407
Figure 7.21	Core Model 1b Elaborated with Involvement Characteristics	407

## LIST OF TABLES

Table 2.1	Three Typologies of Factors Found to Influence the Use of Evaluation Information	45
Table 3.1	Theory-Based Evaluation Data Collection Matrix	89
Table 3.2	A Use Example of the Data Collection Matrix for the Evaluation of the Garbage Reduction Program in Taiwan (Chen, Wang, and Lin (1997)	96
Table 3.3	Matrix for Classification of Theory-Based Evaluations	111
Table 3.4	Classification of Theory Use in Evaluation Reports Found to be Predominantly Implementation Theory-Based	128
Table 3.5	Classification of Theory Use in Improving Access to Academic Support in Higher Education Webb, J., Zhang, C., and Sillitoe, J. (1999)	132
Table 3.6	Classification of Theory Use in Evaluation Reports Found to be Primarily Causative Theory-Based	144
Table 3.7	Classification of Theory Use in Evaluation Reports Found to be Neither Causative or Implementation Program Theory-Based	147
Table 3.8	Classification Matrix Table with Classifications of all Twenty-one Studies.	153
Table 3.9	Program Theory Use Score Guide for Case Study Selection	156
Table 3.10	Classification of Theory-Based Evaluation Reports	158
Table 4.1	Number of Unique People Interviewed by Program	170
Table 4.2	Number of Interviews Conducted	170
Table 4.3	Descriptions of Study Interview Schedules	178
Table 5.1	Perceived Characteristics of the Environment of the Organisation Scale. Item Analysis Results.	207
Table 5.2	Perceived Characteristics of the Leadership Practice of the Organisation Scale. Item Analysis Results.	209
Table 5.3	Perceived Characteristics of the Personal Practice Scale. Item Analysis Results.	214
Table 5.4	Perceived Characteristics of the Organisation's Use of Information. Item Analysis Results.	216
Table 5.5	Perceived Characteristics of the Organisation's Teamwork Scale. Item Analysis Results.	219
Table 5.6	Perceived Characteristics of the Stakeholder Commitment to the Program. Item Analysis Results.	226

Table 5.7	Perceived Characteristics of the Stakeholder Commitment to the Evaluation Study. Item Analysis Results.	228
Table 5.8	Perceived Characteristics of the Stakeholder Involvement in the Evaluation Process Scale. Item Analysis Results.	232
Table 5.9	Perceived Stakeholder Characteristics: Education and Training in Social Research and Evaluation Scale	234
Table 5.10	Perceived Opportunity of Stakeholders to Use Information Scale	236
Table 5.11	Perceived Stakeholder Commitment to the Organisation Scale	239
Table 5.12	Evaluator Role Ratings	244
Table 5.13	Involvement Rating Calculation and Final Scores	250
Table 5.14	Evaluation Involvement by Person	252
Table 5.15	Evaluation Involvement Total Score by Person	254
Table 5.16	Evaluator Perception of Evaluator Practice Scale. Item Analysis Results.	262
Table 5.17	Perceived Characteristics of Evaluator Regarding Stakeholder Involvement in the Generic Evaluation Process. Item Analysis Results.	263
Table 5.18	Perceived Process Use Scale. Item Analysis Results.	268
Table 5.19	Stakeholder Perspective of the Evaluation Study Characteristics Scale. Item Analysis Results.	273
Table 5.20	Stakeholders' Perspective of the Influence of the Evaluation Study Findings Scale. Item Analysis Results.	274
Table 5.21	Stakeholders' Perspective of the Evaluation Team Characteristics. Item Analysis Results.	277
Table 5.22	Stakeholders' Perspective of the Evaluation Study Environment Scale. Item Analysis Results.	278
Table 5.23	Stakeholders' Perspective of the Evaluation Study Involvement Scale Items.	281
Table 5.24	Stakeholders' Perspective of the Influence of the Use of Program Theory in the Final Evaluation Report Scale: Item	283
Table 5.25	Stakeholders' Perspective of the Influence of their Involvement in the Program Theory Elaboration Process Scale. Item Analysis Results.	286
Table 5.26	Manner of Evaluation Information Use that are the Focus of this Section	288

Table 5.27	Stakeholders' Perspective of the Use of Evaluation Information- Importance of Use. Item Analysis Results.	293
Table 5.28	Stakeholders' Perspective of the Use of Evaluation Information- Likelihood of Use. Item Analysis Results.	294
Table 5.29	Stakeholders' Perspective of the Use of Evaluation Information- Likelihood of Use Sub-scales. Item Analysis Results.	297
Table 5.30	Stakeholders' Perspective of the Use of Evaluation Information- Importance of Use Sub-scales. Item Analysis Results.	298
Table 5.31	Evaluator Perceived Characteristics of the Evaluation Scale	301
Table 6.1	Theory-Use Classification of the YOHFest Evaluation	317
Table 6.2	Theory-Use Classification of the Pre-Apprenticeship Evaluation	326
Table 6.3	Theory-Use Classification of the Making Consistent Judgments Evaluation	336
Table 6.4	Theory-Use Classification of the Pre-Apprenticeship, YOHFest and MCJ Program Evaluations	338
Table 6.5	Program Theory Use Score Pattern	340
Table 6.6	Program Theory Use Scores for Case Study Evaluations	340
Table 7.1	Goodness of Fit Measures for Measurement Model Latent Variables	355
Table 7.2	Eight Versions of the Core Model	377
Table 7.3	Standardised Regression Weights for Model Version 1	380
Table 7.4	Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Study Findings (Model Version 1)	381
Table 7.5	Standardised Regression Weights for Model Version 2	384
Table 7.6	Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Use of Program Theory in Final Report (Model Version 2)	385
Table 7.7	Goodness of Fit for the Eight Core Models	387
Table 7.8	Regression Weights for Model 1 Elaborated with OLE Variable	392
Table 7.9	Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Study Findings with inclusion of Organisational Learning as a Predictor Variable (Model Version 1)	393
Table 7.10	Regression Weights for Model 2 Elaborated with OLE Variable	395

Table 7.11	Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Use of Program Theory in Final Report with Inclusion of Organisational Learning as a Predictor Variable (Core Model Version 2)	396
Table 7.12	Goodness of Fit Measures for Core Structural Model Versions Elaborated with Organisational Learning Environment	397
Table 7.13	Regression Weights for Model Version 1 Elaborated with Evaluator Characteristics	400
Table 7.14	Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Study Findings with Inclusion of Evaluator Characteristics as a Predictor Variable (Model Version 1)	401
Table 7.15	Regression Weights for Model Version 2 Elaborated with Evaluator Characteristics	403
Table 7.16	Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of the Use of Program Theory in the Final Report with Evaluator Characteristics included as a Predictor Variable (Model Version 2).	404
Table 7.17	Goodness of Fit Measures for Core Structural Model Versions with Evaluator Characteristics Latent Variable Added	405
Table 7.18	Regression Weights for Model Version 1 Elaborated with Expected Study Involvement and Study Involvement (Post)	408
Table 7.19	Direct, Indirect and Total Effects of Latent Predictor Factors on the Influence of Study Findings with inclusion of Involvement Predictor Variables (Expected and Actual) (Model Version 1)	409
Table 7.20	Goodness of Fit Measures for Elaborated Model Version 1a and 1b with Involvement Variables (Expected and Actual) Added	410

# **CHAPTER I**

## **INTRODUCTION**

The use of program evaluation information has concerned program evaluators and stakeholders for decades, because information resulting from an evaluation study is little utilised. The theory-based evaluation approach, first promoted by Suchman (1967) four decades ago, is one of a number of endeavours by program evaluators to enhance program-related knowledge and decision-making. There have been strong exhortations over the past 20 years for evaluations to be conducted in accordance with a program's theory, although few descriptions of the practice have appeared in the literature on a regular basis. It is surprising that in spite of the support for, and apparent popularity of, the program theory-based approach in undertaking an evaluation, few examples can be found in evaluation literature. Rogers, Petrosino, Huebner and Hacsí (2000) note that theory-based evaluations of small or local projects are frequently not published nor distributed widely except as conference papers.

It is also surprising, and of greater concern, that little empirical research has been undertaken to investigate the influence of theory-based program evaluation on the utilisation of the information from an evaluation study. Clearly, advocacy for a practice should be supported by relevant empirical evidence. The aim of this research study is to address the shortcoming in empirical research by gathering information that will provide insight into the extent to which program theory use has been used as the basis for an evaluation study, and the nature and extent of the utilisation of the information resulting from the study.

Identifying what comprises a theory-based approach to undertaking an evaluation study is important to understanding the central theme of this investigation. One of the primary differences between the modus operandi of an evaluation underpinned by the use of a program theory and that of other evaluation approaches is the basis upon which decisions about the evaluation



study are made, including the focus of the instruments developed for data collection and the method used to conduct the study. If those advocating theory-based evaluation are correct, use of program theory in an evaluation should have an impact on the utilisation of the evaluation and the information it yields. The literature underpinning the context of the history and development of earlier evaluation approaches more commonly adopted, as well as the theory-based evaluation approach, is summarised in Chapter II.

## MODERN PROGRAM EVALUATION

Modern social program evaluation began its development in association with social programs developed to address needs in education, health, criminal justice and housing initiated in the United States under President Kennedy and further developed under Presidents Johnson and Nixon. US federal legislation mandating the undertaking of program evaluations (and the appropriation of funding for it) was a catalyst in the establishment of professional evaluation (Shadish et al., 1991).

Those involved in laying the foundations of modern program evaluation as a discipline, relied heavily upon traditional quantitative social scientific research methods in an attempt to establish evaluation as a new science for assessing program worthiness (Chen, 1990b). Optimism that social programs would achieve their objectives encouraged evaluator to focus on the input-output link directly, and attention to program processes was rare (Chen, 1990b; Conrad and Miller, 1987). Therefore, the design of many early evaluations of social programs considered the program as a closed unit, examining the focus of the program prior to implementation (i.e. population, event) and the outputs to determine the extent of the program's success, yet failing to examine the critical processes or causal mechanisms linking the inputs and the outputs. For example, measurements of the program effects were collected, but information regarding implementation procedures, program processes, interactions between those implementing the program and the target group, and treatment selection often were not connected (Rossi and Wright, 1986). This type of evaluation has come to be known as 'black-box' evaluation (Rossi and Wright,

1986; Lipsey, 1997). The processes of the program, which are the content of the black-box, generally are ignored in favour of measuring the inputs and the outputs of the program (Rodriguez and Mead, 1997). Such evaluations usually provide little or no evidence as to why or how a program works. As Weiss (1972) has argued, “Knowing only program outcomes, even if we know them with irreproachable validity, does not tell us enough to inform program improvement or policy revision” (p. 51).

#### PROGRAM THEORY-BASED EVALUATION

In 1967 Suchman argued that “The evaluation researcher who approaches his task in the spirit of testing some theoretical proposition rather than a set of administrative practices will in the long run make the most significant contribution to program development” (p. 75). Even when the theory or ‘rationale’ underlying a program is not tested explicitly, a critique of this rationale can also be valuable, as Stake (1967) pointed out. According to Palumbo and Petersen (1994) this is because when the theory underlying a program is considered, insight is available into the intention and implications of the program.

Rogers et al. (2000) broadly define program theory-based evaluation as consisting of “an explicit theory or model of how the program causes the intended or observed outcomes and an evaluation that is at least partly guided by this model” (p. 5). They consider theory-based evaluation as an approach in which the program processes are elaborated for investigation in accordance with the theory of the program. A fuller consideration of what is understood to be a program theory will be presented later in this dissertation. It is important however, to recognise a major distinction between the causative theory of a program and a program’s implementation theory.

Program *implementation theory* addresses the variables affecting how a program is implemented, whereas a program’s *causative theory* identifies the causal links between the predictor variables, which govern a program’s application, and its intended effects (Scheirer, 1987). Although they are

interrelated, it is essential to identify and distinguish between these two types of program theory because analysing both in an evaluation provides a fuller picture. Identifying the intervening variables (*program causative theory*) and carefully considering their effect in the implementation of the program (*program implementation theory*) can provide policy-makers and program developers with useful information regarding a program's operation. It is difficult to differentiate program causative theory failure from program implementation theory failure unless it can be shown that a program was faithfully implemented in accordance with the intended program implementation theory. As Bickman (1987) has pointed out, there are numerous examples in the literature of interventions or programs found to have no effect and hence the program causative theory was discredited when in fact the intervention was not implemented as intended, and so the program causative theory which underpinned the program was not 'tested' and, in fact, may be sound.

Although the concept of theory-based evaluation has been in use for forty years, it is surprising that program evaluations which account for a program's theory have appeared in the literature only in the past twenty years. In general, it was evaluation researchers' consideration of their field of practice in the early 1980s that provided the momentum for theory-based evaluation in the mid 1980s (Rossi, 1990). The growing popularity of theory-based evaluation since this time can be seen in the publication of three issues of 'New Directions for Program Evaluation' which have focused on theory-based evaluation strategies (Bickman, 1987, 1996; Rogers et al., 2000). Many evaluators have acknowledged the usefulness of program theory in the design of evaluations of program processes and outcomes including Chen (2003, 2004), McLaughlin and Jordan (2004), and Rogers et al. (2000). In spite of the views of some credible critics, such as Scriven (1994), significant interest in program theory-based evaluation continues, and it is frequently undertaken in some form.

## EVALUATION UTILISATION

In a paper presented to the American Sociological Association in 1966 it was stated that the most common complaint of evaluators of that time was that the findings of the program evaluations they undertook were not utilised (Weiss, 1972c). A decade later, in response to continued rising concern over the lack of evaluation information utilisation, the Joint Committee on Standards for Educational Evaluation (1981) undertook a five-year project that resulted in the publication of a set of standards calling for evaluations to have four key features: utility, feasibility, propriety and accuracy.

The Joint Committee's rationale for the ordering of these features, with *utility* leading, is essentially that an evaluation should not be done at all if there is no prospect for its being useful to some audience (Stufflebeam, 1980). The conduct of a quality program evaluation intent on alleviating a difficult social problem, is a task undertaken primarily to inform and improve the social intervention under consideration. If the information from such an evaluation is ignored, the time, energy and vision dedicated to the evaluative effort have been squandered (Shadish et al., 1991).

The question of what has to be done to get results that are appropriately and meaningfully used has preoccupied numerous evaluation researchers for some time. This concern with the perceived under-utilisation of evaluation information was also the catalyst for the consideration of factors that might influence evaluation information utilisation. Weiss (1972c) began the trend by presenting a variety of factors that might account for under-utilisation in relation to what is now termed *instrumental use*. Evaluative researchers were looking for predictors of evaluation information use based on the belief that, if evaluators were cognisant of the factors that shape the potential of an evaluation's impact, these factors could then be incorporated into the design and implementation of an evaluation study (Seigel and Tuckel, 1985). Numerous authors have contributed to the growing program evaluation utilisation literature and to the slow-paced, yet balanced, developments in the understanding of the use of program evaluation information (e.g. Alkin et al.,

1979; Cousins and Leithwood, 1986, 1993; Kirkhart, 2000; Leviton and Hughes 1981; Shadish, Cook and Leviton, 1991).

#### SIGNIFICANCE OF THE STUDY

The evolution of program theory-based evaluation is yet another step in the endeavour of evaluators trying to increase the appropriate use of program evaluation information. The most common impetus for conducting evaluations using a program's underlying rationale or theory to understand the impact of the program, is the promise of increased explanatory power (Smith, 1994; Weiss, 1997a).

Advocates of program theory-based evaluations see this approach to be the solution to a variety of problems in evaluation (Rogers et al., 2000). The ability of theory-based evaluations to expand understanding of how a program works, in addition to providing information regarding whether a program works is claimed to lead to an increase in the usefulness of program evaluation information in terms of program and policy development (Chen, 1990b, 2004; Pawson and Tilley, 1996). Therefore, it is surprising that there are relatively few examples of program theory-based evaluations to be found in the formal evaluation literature. Rogers et al. (2000) suspect that program theory-based evaluations of small projects are not published nor widely distributed, but rather are only documented in an evaluation report or are the focus of conference papers. Furthermore, of the examples available, most fail to describe any systematic testing of the causal program theory (Rogers et al., 2000). It is the aim of this study to contribute to rectifying this lack of information by undertaking a study focused on the impact of theory-based evaluation practice on the use of evaluation information.

## PURPOSE OF THE STUDY

The effect of the theory-based evaluation approach on the use of the information of an evaluation is the primary interest of this study. In 1977 Straton raised questions about research on the evaluation process, suggesting research was needed with regards to: the identification of the kinds of questions the evaluation is to address, or *delineating* the evaluation focus; information collection processes adopted in undertaking an evaluation, or the *obtaining* of evaluation information; the most effective way of *providing* evaluation information to the audiences; and methods of enhancing the *utilising* of the evaluation's information. Evaluation utilisation literature has commonly focused on the impact of a range of factors on the use of evaluation information (e.g. Alkin, 1985; Cousins and Leithwood, 1986, 1993; Hudson-Mabbs, 1993; Leviton and Hughes, 1981).

This study adds to the understanding of the use of program theory in the conduct of an evaluation study and to the role of the factors examined. In particular, this study investigates three research questions:

- i. What is the influence of program theory on the use of evaluation information?
- ii. Which factors have the greatest impact on the use of evaluation information?
- iii. How do these factors interact with each other to affect use?

In order to investigate the effect of program theory use (and other factors) on evaluations and their outcomes, a model has been developed which includes factors considered influential to the use of evaluation information. Ten factors are examined and their influence on four evaluation information utilisation outcome variables is investigated. The logic represented in the model underpins the focus of this study and frames its findings.

## THE STUDY METHOD

The main focus of this study is on the response of the program stakeholders to an evaluation study and the evaluation's information. Three cases were selected for investigation in accordance with predetermined requirements regarding the degree and type of program theory use in the evaluation study. The premise of this approach is that the nature and extent of the program stakeholders' response reflects the use of the information. The variations in theory use among the three evaluation studies were expected to enable the analysis of the effect of theory use, by degree and type, on the use of evaluation information. One evaluation study was selected for its strong use of program implementation theory, another for its strong use of program causative theory, and the third study was selected for its apparent disregard of program theory. Therefore, the initial undertaking in selecting an appropriate sample of subjects for the study was to identify suitable evaluation studies which could be used as 'cases' to be investigated. The stakeholders of the evaluated programs comprise the primary subject group of this study. The members of other subject groups from whom information was obtained include the evaluators of the programs and representatives of the organisations responsible for the delivery of the programs.

Each study was followed longitudinally with interviews undertaken at key points in the evaluation process: at the beginning, midway through the conduct of the study, and following the release of the final report. This concurrent longitudinal study of three evaluation studies of programs as they occurred was undertaken with a view to identifying the use of the information produced by each evaluation, and determining the characteristics which influenced the use of the evaluation and its information. The evaluations selected were in each case undertaken by evaluation consultants independent of the sponsoring organisations of the programs. Each evaluation study had as its focus projects with disparate target populations and stakeholders to the other two evaluation studies selected as cases. Since the program stakeholders form the primary subject group of the study, it was important that each evaluation had a distinct stakeholder group so that the extent of evaluation information use could be

determined. If the stakeholder groups of the theory-based or comparison evaluations had overlapped, confounding of the study findings and bias could have been a concern. Due to the complex nature of program evaluation information use, in addition to interviews with stakeholders and evaluators of the program evaluations, document reviews, observation and informal communications were used to collect information which provided valuable insight into the quantitative findings of the study.

## ORGANISATION OF THE DISSERTATION

Chapter II of this dissertation provides a review of significant literature relevant to the study. The chapter begins by broadly considering the nature of program evaluation before leading into the development of the theory-based evaluation approach and focusing on issues surrounding the conceptualisation, implementation and use of theory-based evaluation. The remaining sections of the chapter concentrate on the use of evaluation information. In particular, the logic underpinning the conceptualisation of evaluation utilisation is considered and research concerned with the use of evaluation information is reviewed.

Chapter III focuses on the development of a conceptual model to be investigated and contains three sections. The conceptual framework that forms the basis of this study is presented in the first part of the chapter, describing a model of program theory development and detailing the logic underpinning its conceptualisation. The second section focuses on the form of the model to be tested in this study and includes a description of the predictor and outcome variables of the model. In addition, further elaboration of the model is presented, including all of the variables constructed, to reflect as closely as possible the conceptual model. In the final section of Chapter III a matrix to classify program theory use in evaluation studies by source, type, degree and level of contextualisation is introduced. The classification of twenty-one evaluation studies in accordance with the matrix is then detailed.

The study methods are presented in Chapter IV. The initial focus of the chapter is on the study intent and design. The selection of the three cases, the



criteria used the selection of the sources of information and the data collection instruments of the study are then described. The sources of information include a review of pertinent documents, structured interviews, observation and informal communications. A summary of each of the interview schedules used in the study is then presented, with a particular focus on the measurement of the variables in the model. The procedure of the study is discussed in the final section of the chapter.

The six interview schedules developed to collect the qualitative and quantitative data for the study are the focus of Chapter V. The chapter first discusses how the latent variables for the structural equation model analysis were operationalised. Each interview schedule is then considered in turn. For each schedule the logic underpinning the development of each instrument is explained and the structure of each schedule presented. A description of the items in each scale and a scale and item analysis for each scale and a summary of the changes made as a result of the scale analysis process is included. Finally, a description of the analysis process of the scales is provided.

Chapter VI provides a description of each case study. Each program is described, identifying the primary stakeholders and, where appropriate, the causative and implementation program theories underpinning the programs. Finally, the evaluation study of each program is described and discussed and the use of each program's theory classified into the Classification Matrix developed for this investigation.

The findings of the study are given in Chapter VII. The structural equation modelling process is described, including the use of Analysis of Moment Structures (AMOS version 7) to test the measurement model and the structural model. Several versions of the study model were developed to explore the influence of program theory in an evaluation on evaluation information use. The fit statistics of the models are determined and the extent to which each model fits the data is reported.

A discussion of the findings is presented in Chapter VIII. The implications of the models found to fit the data in Chapter VII are considered in the light of significant regression estimate values, effect values and multiple squared correlation values. The latent predictor variables of each model are reviewed with consideration given to their effect on other latent variables and on the outcome variables. Qualitative data gathered in the study is used to provide contextual insight to the study's quantitative findings.

The final chapter, Chapter IX, begins with a brief summary of the study and its conclusions. The purpose, conceptual underpinnings, focus, design and methods of the study are outlined. The implications for further research are considered. The chapter concludes with a discussion of the implications of the study for evaluation practice.

## CHAPTER II

### LITERATURE REVIEW

The benefit of the use of program theory in undertaking an evaluation is a frequent subject of focus of the evaluation literature (Bickman, 1987; Chen, 2004; Weiss, 1998). Subsequently, few exemplary cases of theory-based evaluation are available in the formal evaluation literature. Rogers et al. (2000) found many of the evaluations they were able to locate in a search for theory-based evaluations conducted in the 1995-1999 time frame used program theory in a limited and specific manner (e.g. to plan an evaluation). Of those they identified, they found few which used the program theory extensively to guide the evaluation.

The first step of this literature review is to consider the background of program evaluation within which the program theory-based evaluation practice has developed with a view to understanding, at least conceptually, what is meant by a program theory-based evaluation. Literature focused on defining theory-based, or theory-driven evaluations undertaken is then examined. Finally, a review of literature pertinent to the impact of evaluation information, including theory-based evaluations, is presented.

This chapter is presented in three sections. The first section details the emergence of theory-based evaluation as an approach to undertaking a program evaluation study, considering the physical and social influences contributing to its emergence. The second section addresses program theory, a concept around which there tends to be some confusion, due primarily to a lack of formalized guidelines for determining what is and what is not a program theory. The more common conceptualisations of program theory found in the literature are presented. This is followed by a consideration of three sub-theories that contribute to the aggregate concept of program theory. These sub-theories are a program's *prescriptive theory*, *causative theory* and *implementation theory*. The focus on the underlying mechanisms responsible for a program's outcomes, in other words the program's causative theory, is a

unique facet of theory-based evaluation. As a result this section of the chapter delineates causative theory and tends to reflect the abundance of work in this area in its consideration of related aspects including the nature of the causative theory adopted for an evaluation, the role of the evaluator in the evaluative process, and the role of context.

The third section of the chapter focuses on evaluation information use. The evolution of the definition of evaluation information utilisation is traced and commonly accepted evaluation use types are defined before the use of evaluation information is considered in the program theory development model. The influence of theory-based evaluation information on the model of information use is then reflected upon. Finally, the factors influencing the use of evaluation information are presented.

## EMERGENCE OF THE THEORY-BASED EVALUATION APPROACH

### **EARLY PROGRAM EVALUATION INFLUENCES**

The ‘positivism orthodoxy’, conceptualised with the natural sciences as a reference point, has had a significant influence on social research. In general, the positivism perspective recognises only two kinds of knowledge as meaningful: the empirical substantiated by science and the logical confirmed by mathematics. This philosophy is traceable to the empiricism philosophy introduced in the works of Frances Bacon, Thomas Hobbs, David Hume and John Stuart Mill, and does not differentiate between the epistemology of the natural sciences and the social sciences (Hughes, 1990). Heavily influenced by this early tradition, social scientists commonly define social theory as a set of interrelated propositions that explain and predict a phenomenon (Kerlinger, 1979; Lave and March, 1993), so as to objectively study events as they are without extending propositions regarding the way these events should be (Chen, 1990b). The theories conceptualised according to this view are traditionally descriptive, empirical, hypothetico-deductive models (Argyris, Putman and Smith, 1985), intent on explaining causation in successionist

terms (Harre, 1972) through consideration of external observable events (Pawson and Tilley, 1994).

The 1960s saw the emergence of modern social program evaluation. Social programs in education, income maintenance, housing, health and criminal justice were initiated in the United States under the Kennedy administration and subsequently under Johnson, with substantial funds being appropriated to support them (Shadish, Cook and Leviton, 1991a). The unwavering conviction of the time was that the social programs which were developed and implemented, enabled through the rapid growth in social welfare spending of the time, would ameliorate social problems quickly (Conrad and Miller, 1987; Pawson and Tilley, 1997; Shadish, 1987). It soon became clear, however, that the efficacy of these programs needed to be determined and monitored, i.e. they needed to be evaluated.

The concepts and principles of modern program evaluation were initially developed during the 1960s. Although parallel developments in the field occurred in Canada, Sweden and West Germany, and in the United States, the decision by Congress to fund some large educational programs and appropriate resources to evaluate those programs was the most obvious catalyst for the boost to program evaluation (Scriven, 1991). Following on the requirement from Congress that educators evaluate their work, it was the allocation of millions of dollars by federal, state and local governments that enabled the improvement of educational evaluation practice (Joint Committee on Standards for Educational Evaluation, 1981). These developments were monitored with considerable interest by a number of social researchers in Australia, particularly in education.

With the investment of huge amounts of funds came concerns regarding the accountability of federal fund recipients and the impact of the funded programs. Simply, program evaluation evolved in response to the need for programs to be assessed for the purpose of justifying the resources allocated to them (Cook, Leviton and Shadish, 1985). The initial optimism that the new social programs would work, limited interest in conducting evaluative studies

which included process measures. Preference was given to a ‘methods based’ approach for testing the input-output link directly (Conrad and Miller, 1987) and attention to the program processes was rare (Chen, 1990b).

## **PROGRAM EVALUATION AS EXPERIMENTAL RESEARCH**

Suchman (1967) conceived of evaluation as research, restricting the evaluation process to scientific research methods and practices. His ideas were congruent with those of Donald Campbell who, in his work at Northwestern University in the 1970s, argued strongly for the desirability of the classic randomized, controlled experimental paradigm and helped to establish this approach early in the development of program evaluation (Shadish et al., 1991a).

Donald Campbell’s influential and well-regarded training program on evaluative research conducted at Northwestern University in the 1970s served as an important medium for the entrenchment of the classic experimental paradigm in program evaluation (Finney and Moos, 1992; Pawson and Tilley, 1997; Rossi and Wright, 1986). Many of the early evaluations considered to be of most value were conducted using these designs to determine the effects of an intervention or program in empirical terms. His conceptual and empirical publications of the time, and those of some others, emphasised the importance of rigorous scientific methods in the conduct of good quality program evaluation studies (Cook and Campbell, 1979; Suchman, 1967). Thus, the findings of studies with high internal validity were considered superior to those of less structured quasi-experimental designs as a necessary foundation for practice (Finney and Moos, 1992; Pawson and Tilley, 1997; Rossi and Wright, 1986).

For this school of thought, program outcomes were almost exclusively relied upon to determine the effects of a program (Shadish et al., 1991a). Marginal program change was ignored for the most part, with evaluation studies considering local program practice responding to local circumstances being seen as less interesting (Shadish et al., 1991a). Alternatively, the Stanford Evaluation Consortium directed by Cronbach (Cronbach et al., 1980) in the

1970s raised the level of evaluations significantly in working with the State of California to examine its relationships with local school districts, among other projects, emphasising the importance of local contexts on performance, and the social and political aspects of program evaluation.

### **NON-TRADITIONAL PROGRAM EVALUATION INFLUENCES**

At the same time as Donald Campbell's program was having a significant influence on program evaluation in the United States, a small group of educational researchers in the United Kingdom together with a few colleagues in the United States were beginning to voice their concerns about evaluation practice founded upon the experimental paradigm. In 1972 fourteen such researchers, chosen for their support of 'non-traditional' (Hamilton et al., 1977, p vii) evaluation methods, met at the first 'Cambridge Conference' with the intent of exploring approaches to curriculum evaluation alternative to the currently dominant evaluation practice, and to agree upon guidelines for the further development of the field of evaluation (Hamilton et al., 1977).

Conference members developed an agreed summary of their conclusions which they drafted in the form of a manifesto. The first conclusion was aimed squarely at experimental design-oriented evaluation practice, taking issue with its lack of attention to program processes, claiming it had an over-emphasis on "psychometrically measurable changes in student behaviour" (Hamilton et al. 1977, p. viii) to represent complex changes and denouncing its disregard for problematic communication between researchers and others. In general, the manifesto authors agreed that traditional psychometric experimental methods-based evaluation practice did not provide sufficiently useful information regarding the complex problems present in the programs they were investigating, resulting in little information useful to the decision-making process. Alternatively, they supported more flexible responsive, illuminative and democratic evaluation practices incorporating more observational data and reported in terms accessible by non-research-oriented audiences. The alternative evaluative paradigm which they advocated, they noted, fits more comfortably with the fields of social anthropology, psychiatry and sociology,

than with the experimental mental-testing traditions of psychology adopted in early educational evaluation research (Parlett and Hamilton, 1977).

In 1972, the same year as the first Cambridge Conference, Carol Weiss published an article in which she proposed the use of the 'program model' in undertaking an evaluation study, making hers the earliest reference to what she now terms "theory-based evaluation" (Weiss, 1997b, p. 41). The program model represented in the article incorporates program processes, as intended, constructed through the identification of the steps and mechanisms by which a program is proposed to operate. This evaluation approach was similar to that advocated by the Cambridge Conference manifesto in that it supported the consideration of complex program processes in an evaluative undertaking, yet it was dissimilar to a majority of the program evaluations conducted.

Other evaluative researchers of that time advocated alternative perspectives concerned with the quality of the program implementation and causal processes mediating program influence (Shadish et al., 1991a). These perspectives included: Bloom, Hastings and Madaus' (1970) use of student testing in evaluation practice; Popham's (1973) advocacy of educational evaluation anchored to behavioural objectives; Cronbach's (1963) use of instructional development in evaluation; Stufflebeam et al.'s (1971) consideration of decision-making at the administrative level; Scriven's (1967) orientation of evaluation as a service to consumers; Glaser and Strauss' (1967) focus on grounded theory; and Stake's (1967) conceptualisation of evaluation as two basic and necessary acts essential to understanding educational programs, description and judgment. Other later evaluation theorists made further contributions to evaluation, such as Guba and Lincoln (1981), who included qualitative methods as an alternative to the traditional quantitative approaches, Cronbach and Associates (1980) who considered the evaluator's role to be that of an educator, and Greene (1987) who advocated including the stakeholder as a participant in the evaluation.

Following Weiss's (1972a) early contribution about theory-based evaluation, Fitz-Gibbon and Morris (1972) added to the early work on this theme. They



also advocated the use of theory-based evaluation as a mechanism for generating further program knowledge and presented the first detailed description of theory-based evaluation, stating that:

A theory-based evaluation of a program is one in which the selection of program features to evaluate is determined by an explicit conceptualisation of the program in terms of a theory which attempts to explain how the program produces the desired effects (p. 177).

In summary, the early establishment of the classic randomized experimental paradigm in program evaluation is undeniable. However, the work of many influential evaluative researchers reflects a growing realization that this approach often did not provide sufficiently insightful program information to aid decision-making. Discrepancies were more and more frequently identified between the information resulting from an evaluation underpinned by the experimental paradigm and information necessary for insightful decision-making. Furthermore, it was increasingly found that this approach often was inappropriate for the circumstances surrounding many of the programs which needed to be evaluated. In terms of evaluating social programs designed and implemented with the goal of having an impact upon people's lives, numerous ethical and legal issues were increasingly seen to arise when considering the conduct of a randomized, controlled experiment. Finally, experience had shown that rigorous field experiments were expensive, time-consuming, and difficult to conduct (Rossi, 1990; Rossi and Wright, 1986; Shadish et al., 1991a).

#### **PROGRAM EVALUATION CONSIDERATION OF 'NO EFFECTS' FINDINGS**

Beginning in 1980, drastic social program cutbacks in the United States and the United Kingdom, and a general policy of unwillingness to meet rising costs of programs, drastically reduced the call for evaluation research (Fetterman, 1988; McTaggart, Caulley and Kemmis, 1991; Rossi, 1990; Rossi and Wright, 1986; Worthen, 1996). The resultant lull in program evaluation research in these countries enabled two important developments according to

Rossi (1990). First, gaps in information about the condition of society and of program effects allowed policy-makers an opportunity to appreciate the value of such information. Hence, the appreciation of, and desire for, social research and evaluation increased. Secondly, this period of reduced evaluation activity allowed energies normally directed towards the conduct of evaluations to be redirected towards the concepts, principles and practices of program evaluation. Time was available to reflect on what had been learned in the previous two decades of evaluation practice. Of uppermost concern was the finding that many of the social programs evaluated apparently had little or no effect, stimulated by the publication of Martinson's (1974) often cited 'What works? Questions and answers about prison reform' (Chen and Rossi, 1981; Pawson and Tilley, 1997).

Critical consideration of the 'no effects' finding inspired a general acknowledgment that it was unreasonable to assume that most programs were so poorly designed they had no impact (Chen and Rossi, 1981). A rethinking of optimistic program expectations occurred (Conrad and Miller, 1987; Rossi and Wright, 1986; Shadish, 1987) and the resulting reduced expectations allowed the merits of the incremental changes induced by social programs to be considered. Over time the expectation of quick radical social problem amelioration was no longer held as the realistic goal of social programs (Conrad and Miller, 1987; Shadish, 1987). Furthermore, it was acknowledged that program evaluation research designs to that time had at least been adequate to detect some program impact where it existed. Thoughtful concentration on this dilemma led evaluation researchers to suspect that the problem lay in the discrepancy between the program goals delineated by program designers and actual program outcomes (Chen and Rossi, 1981), a notion addressed previously by Robert Stake (1967).

## **THE USE OF PROGRAM THEORY IN PROGRAM EVALUATION RECONSIDERED**

In 1981 there appeared the first in a series of publications by Chen and Rossi (1981) advocating "theory-driven" (p. 38) evaluations, resting their work on the assumptions that all programs have some effect, and that expected program

effects may be anticipated through a union of *a priori* program knowledge and social science theory. Similar to Weiss's (1972a) theory-based approach, Chen and Rossi's (1981) notion of theory-driven evaluation advocated the use of a program's theory in elucidating the mechanisms and processes to be investigated in the evaluative undertaking. However, the sources of the program theory adopted in the two approaches differ. Chen and Rossi (1981) advocate the adoption of a predominantly social science-based theory (i.e. a generalised theory). On the other hand, Weiss (1972a) presents the program theory as a representation of the processes and mechanisms specific to the program under investigation (i.e. local theory).

A critical objective during the planning and development of a program is to identify the cause of some problem or dysfunction with a view to designing a program or treatment to improve it, treat it or solve it (Rodriguez and Mead, 1997). A traditional evaluation of such a program would consider the program as a closed unit, examining the inputs, and outputs or outcomes to determine the extent of the program's success, yet fail to address the critical processes or causal mechanisms linking the inputs and the outputs. For example, measurements of the effect of the program on the status of those targeted by the program may be collected, but information regarding the effect of the program components (i.e. implementation, program processes, interaction, and treatment selection) are not collected (Rossi and Wright, 1986). This type of evaluation has come to be known as 'black-box' evaluation (Rossi and Wright, 1986; Lipsey, 1993). The processes of the program, which are the content of the black-box, are ignored in favour of measuring and comparing the inputs and the outputs of the program (Rodriguez and Mead, 1997). Such evaluations usually provide little or no enlightenment as to why or how the program works or why it failed. As Weiss (1972a) writes, "Knowing only program outcomes, even if we know them with irreproachable validity, does not tell us enough to inform program improvement or policy revision" (p. 51). Furthermore, black-box evaluations have limited value in terms of the the ability to generalise results (Conrad and Miller, 1987). Alternatively, theorists such as Chen and Rossi (1981), Cronbach (1982), Fitz-Gibbon and Morris (1972), Parlett and Hamilton (1972), Stake (1967) and Weiss (1972a) advocate evaluation aimed

at describing and explaining the program processes for the purpose of improving the program under investigation (Cook, Leviton and Shadish, 1985).

As Chen writes, an evaluation not driven by theory is “a set of predetermined research steps that are uniformly and mechanically applied to various programs without concern for the theoretical implications of program content, setting, participants, implementing organisations, and so on” (Chen, 1990a, p. 18). When the theory underlying a program is considered, insight is available into the purposes and implications of the program (Palumbo and Petersen, 1994).

In general, it was the evaluation researchers’ thoughtful consideration of their field of practice in the early 1980s that fuelled the gain in momentum of theory-based evaluation in the mid 1980s (Rossi, 1990). By the late 1980s and early 1990s a series of new publications addressing the use of program theory in the undertaking of evaluations appeared, including three issues of *New Directions for Program Evaluation* which focused on theory-based evaluation strategies.

Changes in evaluation practice has been slow to follow, however. The strong association of evaluation with the traditional experimental designs model or at least quasi-experimental designs model, particularly in the minds of those who commission and fund evaluation studies, continues today to be an impediment to theory-based evaluation. Evaluations yielding findings regarding a program’s main effects are often given priority over evaluations focused on explaining the program or its processes (Reynolds, 1998). When Lipsey, Crosse, Dunkle and Pollard (1985) examined a systematic sample of published program evaluation studies and coded in detail characteristics of these evaluations, they found that less than 30 per cent of these studies advanced any theoretical propositions addressing the causal processes of the evaluated program. Only nine per cent of the studies they considered presented “integrated *a priori* theory within which the specific formulation of program elements, rationale and causal process was embedded” (p. 318). Currently,

many evaluators acknowledge the usefulness of program theory in the design of program process and outcome evaluations (Bickman, 1987; Chen, 1990b, 2004; Funnell, 1997; Pawson and Tilley, 1997; Rogers et al., 2000; Weiss, 1997a, 1997b; Worthen, 1996). In the last decade or so the theory-based approach to program evaluation has been more widely used in a variety of areas (Chen, 2003).

### **CRITICISMS OF THEORY-BASED EVALUATION**

Scriven (1994) is probably the most notable critic of theory-based evaluation. He maintains that an evaluation aimed at determining the worth or value of a program does not need to address a theory regarding either the relationships linking the program components or the internal operations of the program. His concern is that a program theory attempting to address this level of complexity may not be accurate. Furthermore, he queries the means by which such a program theory may be ascertained. He does, however, concede that a theory about how the program operates from an external viewpoint may be useful in undertaking an evaluation. He calls this a “grey box” (p. 77) evaluation. Primarily, Scriven (1994) believes that the principle concern of evaluation is to evaluate; “anything else is icing on the cake” (p. 76). He advocates sound entrenchment of logical evaluation and is disinclined to include “risky” undertakings such as program explanation or recommendations as evaluation protocol.

Stufflebeam (2001), although he believes endeavours to model a program can be useful in terms of identifying measurement variables, suggests any attempts to elaborate a program theory, if a tenable program theory does not already exist, is likely to be problematic. He suggests, for example, that an evaluator might elaborate a poor version of the program theory, the evaluation process might become stuck in the program theory elaboration process, or the program theory accepted as valid early in the evaluation undertaking may be found later in the study to have evolved thereby impeding the evaluation.

Patton (1989) also writes of concerns that information from evaluations conducted in accordance with a theory, primarily derived from a social science theory as initially proposed by Chen and Rossi (1987), may be too esoteric and abstract for stakeholders to make much use of. Furthermore, he is concerned that the level of research sophistication required by such an evaluation may alienate stakeholders from the evaluation process.

Nonetheless, in spite of even the most credible of critics, such as Scriven and Stufflebeam, significant interest in program theory-based evaluation continues. For the advancement of program evaluation as a discipline, the adoption of broader perspectives, such as the theory-based approach, is paramount (Chen, 1994a; Pawson and Tilley, 1997).

#### PROGRAM THEORY

There are a number of different conceptions of program theory. In fact, there is a lack of specific formalized guidelines for determining what is and is not program theory, which encourages practitioners to attach their own meanings to the same words (Weiss, 1997b).

As with the theories adopted in general by social scientists, most definitions of program theory are primarily descriptive or explanatory in nature (Bickman, 1987; Reynolds and Walberg, 1990; Weiss, 1997b), and generally represent a form of causative theory (Chen, 1990b; Reynolds 1998). Most commonly program theory is defined as a logical and credible model of how and why a program works (Bickman, 1987; Reynolds and Walberg, 1990). This is an apparently straightforward and uncomplicated explanation for a rather complex entity. Furthermore, under the umbrella of this undemanding definition there is scope for the coexistence and acceptance of a wide range of meanings for the term 'program theory'.

A review of the relevant literature encourages the differentiation of program theory into various contributing sub-theories with the intent of clarifying what is meant by program theory. A lack of specification of particular

characteristics, such as the influence of values (Conrad and Miller, 1987; Hare, 1983), the origins of program theory (Patton, 1996), the influence of the evaluator (Reynolds, 1998), the distinction within program theory of causative theory and implementation theory (Scheirer, 1987; Weiss, 1997b), and the incorporation of context (Pawson and Tilley, 1997) are some of the current concerns of evaluative researchers with regard to program theory development. Furthermore, a number of authors differentiate between program theories on the basis of their magnitude (Shadish, 1989; Cook, Leviton and Shadish, 1985). For instance, Lipsey (1993) defines: 'small theory' as how and why a program works; and 'large' or 'grand theory' as an academic social science theory.

A logical first step in pinning down the theory upon which a program operates is to look to the program itself. The problem with this seemingly sensible approach is that program developers and implementers generally operate on the basis of a few basic assumptions, often implicit and rarely articulated in terms of a causal network. The task of making sense of a program's theory of operation then falls squarely on the shoulders of evaluators (Bickman, 1987; Chen, 1990b; Patton, 1989; Shadish, 1987; Stufflebeam, 2001; Weiss, 1997b). In undertaking a search for theory-based evaluations published in the periodical literature, Weiss and Birckmayer (2000) found that most theory-based evaluation studies do adopt a reasonable and plausible logic which represents the basis upon which the program or intervention has been constructed. They also found that programs in the health promotion and the risk prevention areas tend to follow well developed plans and tend to be built upon clearly explicated theoretical assumptions. As a result, much of the current theory-based evaluation work is taking place in these fields encouraged by the availability of clearly explicated program theories (Weiss, 1997b).

The notion of program theory as representing the integration of the social sciences and social practice, is a union of prescriptive theory and causative (descriptive) theory (Argyris, Putman and Smith, 1985; Chen, 1990b, 2004). Conrad and Miller (1987) define program 'philosophy', a word they have

chosen instead of ‘theory’, as a system of theory or theories and values that “...defines and guides the structure, population, process and outcomes of the program” (p. 21). Their definition includes both causative and prescriptive elements. Chen (1990b, 2004) is in agreement with this view of program theory providing a clear delineation between the prescriptive and causative sub-theories of program theory. He argues that program theory is not only descriptive (causative) in that it identifies the causal mechanisms that link program treatment or intervention, implementation processes and outcomes, but also prescriptive, as it provides the rationale and justification for the program structure and activities. The sections below consider different versions of program theory including prescriptive theory, causative theory, social science theory and stakeholder theory.

### **PRESCRIPTIVE THEORY**

Program theory begins with values (Chen, 1990b; Conrad and Miller, 1987; Pawson and Tilley, 1997). As Parsons writes, “...the main point of reference for analysing the structure of any social system is its value pattern” (cited in Conrad and Miller, 1987, p. 22). These values, guided by affective, personal and moral beliefs are transmitted through symbolic communication (Leslie, Larson and Gorman, 1973) and guide human social behaviour (Chen, 1990b; Conrad and Miller, 1987) through the establishment of socially accepted norms. Norms are the rules and regulations which generally guide the behaviour of a group. They are prescriptive and proscriptive, specifying how people should, should not and must not, behave in different situations (Leslie, Larson and Gorman, 1973).

The prescriptive sub-theory of program theory represents the values of the program, determines what the structure of the program ‘should be’ and rests upon “...unexamined premises, assumptions, customary procedures, and/or prior knowledge and theory” (Chen, 1990b, p. 43), a viewpoint also expressed by Chen and Rossi (1992) and Conrad and Buelow (1990). It is a “practical knowledge, a form of knowing that is traditionally contrasted to theoretical or scientific knowledge” (Argyris, Putman and Smith, 1985, p. 5). It involves



value judgments, is evaluative by nature, and extends propositions regarding the way to do something better or what should be done. It is generally implicit, unexamined and is taken for granted by program developers and other stakeholders (Chen, 1990b).

Until recently, the dominance of the traditional hypothetico-deductive model founded upon natural science has encouraged the minimisation of prescriptive theory and questioned its scientific legitimacy. Theories of this traditional persuasion describe, and seek to explain, regularities among events, are empirically based, and address the causal relationship hypothesized between an independent variable and the dependent variable (Suchman, 1967).

When a social scientist mimics a natural scientist, as was often the case of evaluative researchers of the 1960s and 1970s, the unearthing of fundamental knowledge useful for the solution of social problems tends to be deterred (Argyris, Putman and Smith, 1985; Chen, 1990b; Parlett and Hamilton, 1972). Argyris, Putman and Smith (1985) in their exposition of 'Action Science' claim that knowledge derived from action cannot be wholly represented empirically. In fact, it has been suggested that since human actions are guided by values, prescriptive theory cannot be ignored (Chen, 1990b; Conrad and Miller, 1987; Patton, 1997; Pawson and Tilley, 1997). The meanings and logic of the action must be explored before any information useful for the prediction of behaviour is established (Argyris, Putman and Smith, 1985).

Specific to social program development, values guide the selection of goals to be achieved by a program. Values also guide the choice of the causative theories employed in accomplishing these goals (Conrad and Miller, 1987). In addition, programs are guided by ethical, religious and political beliefs, traditions of practice and ideas about how to effect social processes that may not be clearly perceived (Chen, 1990b).

Cabatoff (1996) provides a useful theory of policy-making. His observations of the factors which influence the evolution of policy which, in turn, initiates program development to address an identified need or problem, are insightful

not only to policy development, but to program theory development. Policy focus determined by the prescriptive theories of those involved, is a major contributing factor in the development of a program with the intent of alleviating a need or providing a solution to a problem. Both Cabatoff (1996) and Cronbach et al. (1980) consider policy evolution as a process that occurs within the “policy-shaping communities” (Cronbach et al., 1980; p. 100) which include individuals from both the public and private sectors. Policy-shaping communities have in common a concern for a particular social agenda that overshadows any geographical or organisational association.

Cabatoff (1996) observes that the processes which lead to the placement of certain issues on the policy-agenda are somewhat obscure and unpredictable. He differentiates between “incremental” and “non-incremental” (p. 36) agenda change. Incremental change refers to a gradual increase in interest in a subject over an extended period of time. Non-incremental agenda change, alternatively, occurs more suddenly and in response to one or more of a multifarious range of events. Examples of influential events include a change of administration and a national disaster. Once a need or problem has risen to the point where it catches the attention of the policy-making community, and an implicit agreement has been reached that the time has arrived to respond to the problem, concrete decisions are likely to be triggered (Cabatoff, 1996).

### **CAUSATIVE THEORY**

Causative theory incorporates two processes, one descriptive and the other explanatory (Suchman, 1967). Most commonly it is empirically based and provides a representation of the causal relationship between the treatment and the outcome and considers not only the intended, but also the unintended, effects of the program. It focuses on the underlying mechanisms which mediate the causal relationship between the treatment and the outcome variables (Chen, 1989b; Chen and Rossi, 1992; Fitz-Gibbon and Morris, 1972; Lipsey and Pollard, 1989). In essence, it is concerned with how the program works and under what conditions (Pawson and Tilley, 1997). Incidentally, Tuckman (1965) suggests that it is the congruence of a relative causative

theory with appropriate values which is a prerequisite for the implementation of a successful program.

The causative theory adopted in social program development emanates primarily from two sources. One source is the stakeholders of the program, and is encapsulated in their views, thoughts and expectations of how the program works, often based on their observations and experience with other relevant programs. The other source is academic social science theory and research (Chen, 1990b; Patton, 1996; Weiss, 1997a). Patton (1996) suggests that these two origins tend to represent the two ends of a spectrum, the practical (stakeholder) and the abstract (social science theory).

Evaluators undertaking theory-based evaluations frequently are faced with the task of teasing out a coherent causative theory for the program. The task may require the articulation of plausible underlying causal mechanisms to account for the outcomes. The most commonly accepted sources for a causative theory are relevant social science theory, explication of stakeholders' implicit theories, and prior research (Chen, 1990b; Lipsey and Pollard, 1989; Patton, 1996; Pawson and Tilley, 1994; Weiss, 1997a; Wholey, 1987). There is much disagreement regarding the optimal approach for the development of causal theory (see for example Chen, 1990b; Patton, 1996; Trochim, 1985; Weiss, 1997a), but most authors do agree that any attempt to develop a causal theory for a program should incorporate propositions from a range of these information sources for balance (Chen, 1990b; Pawson and Tilley, 1996; Pawson and Tilley, 1997; Weiss, 1997a).

Trochim (1985) advocates that a program theory should be consistent with both social science knowledge, and the experience and assumptions of the stakeholders. Conrad and Miller (1987) also provide examples of program theory development incorporating stakeholders' theories and relevant academic social science theory. Finally, it should be noted that theories, even those supported by the data, are never *proved*, only *supported*. It is always possible that an alternative model may provide an equal or better account of a phenomenon (Weiss, 1972a).

## **SOCIAL SCIENCE THEORY**

It has been suggested that programs built on clearly conceptualised theories are very uncommon (Patton, 1989; Weiss, 1997a) while social programs conceptualised according to a particular social science theory are rare (Chen, 1990b; Riggin, 1990; Trochim, 1985). This is suspected to be due to the general lack of well-developed social science theories which has encouraged program developers to rely on a range of sources in the development of program theories such as the results of research, related literature, intuition, observation, formal and informal discussions with colleagues, and the “rational ruminations of the individual expert” (Trochim, 1985, p. 584).

Social programs are rarely conceptualised within a single discipline and although similar theories may be identified in a range of disciplines, they tend to be bound by semantics to their field of origin, making the conceptualisation of a program theory based on social theory traversing a number of disciplines difficult (Bickman, 1989; Riggin, 1990). Nevertheless, acknowledging that social science theory generally is not as specific in detail as theories of the natural sciences, Fitz-Gibbon and Morris (1972) claim that there is enough theory, particularly in the fields of sociology and psychology, upon which to base program theory development.

The use of academic social science theory for the development of program theory is a deductive approach and is advocated by numerous authors (for example Chen and Rossi, 1981; Fitz-Gibbon and Morris, 1972; Weiss, 1997a). Chen and Rossi (1981) advocate a social science approach to program theory development which originates with a concern that evaluators’ uncritical acceptance of key stakeholders’ values may not reflect the reality of the program, but rather a desirable version of the program. The official program goals expounded by the stakeholders may not be the operative goals followed by the program staff. Furthermore, the common sense or hunches upon which stakeholders generally develop their theory of a program’s operation may not address in detail the complex causal processes underlying the program. As

Weiss (1997b) writes, perhaps a little harshly, “... these theories are elementary, simplistic, partial or even outright wrong” (p. 51).

In the opinion of Chen and Rossi (1981), the adoption of a social science theory guided by the expertise and wisdom of an evaluator is a means to elaborate stakeholders’ ideas and shape them into a logical and thorough causative theory detailing program structure, processes and effects. This is preferable to uncritical acceptance of stakeholders’ views. They recommend that stakeholders’ theories at odds with the accepted social science theory be addressed, but always in conjunction with relevant social theory.

The consideration of a relevant social science theory can extend the conceptual domain of a program theory by identifying variables and the inter-relationship among them which may otherwise be overlooked, thereby increasing the validity of the theory (Riggin, 1990). A recognised strength of program evaluation based on a program theory linked to an established social science theory is that its findings can be applied more widely. The findings of evaluations adopting a program theory based upon stakeholders’ tacit theories and assumptions typically are specific to the program under investigation, and as such are bound to the particular circumstances surrounding that expression of the program. On the other hand, the findings of program evaluations conducted in accordance with established social science theory may be generalised back to social science theory development, may contribute to social science knowledge, may enable the prediction of long term consequences (Fitz-Gibbon and Morris, 1996) and, in general, are likely to have a widened scope of usefulness (Chen, 1990b; Fitz-Gibbon and Morris, 1996; Weiss, 1997a). In fact, the underpinning of program theory with social science theory is a fundamental step towards bridging the “...gap between applied and broader theoretical concepts” (Palumbo and Petersen, 1994, p. 160).

Concern has been expressed that the conceptual work required of the evaluator to make a program’s reality fit the “scholarly, academic, abstract” theory may encourage questionable compromises (Scriven, 1994, p. 377). Furthermore,

Patton (1989) has suggested that the theory-based evaluation approach articulated by Chen and Rossi (1981) may be too esoteric for stakeholders who do not have a research background.

## **STAKEHOLDER THEORY**

There appears to be a consensus that a majority of social programs mainly originate from the perceptions, assumptions and tacit theories of stakeholders (Patton, 1996; Wholey, 1987). In keeping with this belief, numerous authors advocate the teasing out of a program's theory from the stakeholder's assumptions, perceptions and tacit theories (e.g. Conrad and Miller, 1987; Patton, 1989, 1997; Pawson and Tilley, 1997; Wholey, 1987).

Wholey (1987) developed his evaluability assessment process in response to his concerns that evaluation to that time did not address important elements of program reality or respond to the information needs of influential stakeholders. A key element in evaluability assessment is the construction of program models from the viewpoint of key stakeholders and interested groups.

Similarly, Patton (1996) advocates an approach to program theory development in which the primary stakeholders are engaged in a process leading to the generation of their conception of the "program's theory of action" (p. 377). In Patton's (1996) view, a program's theory of action includes a list of the program objectives, including implementation objectives and impact objectives, which exhibit interdependent causal links, and is an explicit representation of program stakeholders' conceptions of the program.

Citing the work of Argyris and Schon (Argyris et al., 1985), Patton (1996) distinguished between 'espoused theories', or stakeholders' desired theories (Chen, 1990b), and 'theories-in-use' or the actuality of practice. Patton (1996) advocates the development of a program's causal theory beginning with the explication and testing of stakeholders' espoused theories to elucidate the program theory in reality. In his view, the implicit espoused theories are what

program practitioners believe to be real and therefore may have real consequences for practitioners' actions. Inconsistencies between the espoused theories and theories-in-use represent areas to be investigated (Kemmis and McTaggart, 2000).

Participatory action research is pertinent to the literature addressing explication of stakeholders' theories. Participatory action research evolved in response to the perceived imposition of research and policy agendas on local or community groups by centralised, inaccessible agencies. It is a non-traditional form of research adopted where research participants take an active role in constructing their social reality and, in turn, critically reflect on that reality with a view to improvement (Kemmis and McTaggart, 2000).

Participant researchers must switch between two viewpoints of the program. One viewpoint is an insider's consideration of the participants' perceptions, practices and context. The second is the viewpoint of an outsider, enabling an altered understanding of the participants' perceptions, practices and context. Traversing between the two viewpoints, it has been suggested, creates a less subjective, more balanced view of a program, enabling the explication of a theory of program practice (Kemmis and McTaggart, 2000).

The challenge for a researcher facilitating participatory action research is to enable participants to reflect upon their practice to discover and express in shared terms the tacit knowledge embedded in their practice (Argyris, Putnam, and Smith, 1985). Opportunities for researcher influence in the process are a concern that is considered later in this dissertation.

It is useful to note that because of the practical nature of evaluation, the widely held belief that theory-based evaluation conflicts with key stakeholder concerns remains an impediment to the widespread conduct of evaluations based on a program's theory (Reynolds, 1998). Even Chen and Rossi (1989) acknowledge the possibility that stakeholders' theories may become those most commonly adopted in a theory-based evaluation. They note that although these theories may not necessarily be accurate, they do concede that it may not

be possible to ignore the contribution of implicit theory to policy development and program design. Furthermore, Chen and Rossi (1981) are not completely opposed to stakeholders' views informing program theory developed for a theory-based evaluation and in fact advocate the inclusion of both perspectives. Nevertheless, they do not actually advance the concept so far as to provide guidelines for integrating the two perspectives to form a single program theory (Chen, 1992).

Finally, Shadish (1987) believes that adopting the ideology of stakeholders as a program's theory is unlikely to lead to the identification of fundamental factors impinging upon the program and its effects. His concern is that these beliefs, values and norms represent, to some extent, the problem addressed by the program. The socialisation of a dominant majority to accept a political and economic ideology deters them from the approval of a program that is not closely in tune with prevailing beliefs.

#### THE ROLE OF CONTEXT

At this point a consideration of the role of a program's context in the development of its program theory is appropriate. Early work in the field of program evaluation addresses the influence of context on programs. Suchman (1967) presents the conceptualisation of a program's causal theory as a two part process. First, the relationship between the independent and dependent variables, representing the program and the expected outcome respectively, is tested for legitimate causality, allowing for the description of the relationship. Second, once the causal relationship is verified as genuine, the process and conditions under which the program produces the desired effect are explained. Similarly, in the same year Stake (1967), focusing on what could be considered the important data to be gathered by an evaluator about an educational program, specifies data describing the antecedents, transactions and outcomes of the program. He defines antecedent data as "any condition existing prior to teaching and learning which may relate to outcome" (p. 5), for example the context and the inputs. Furthermore, he writes of 'antecedent' conditions ranging from the status of the individual program client to the



investment of community resources. Then, a few years later, Weiss (1972a) wrote of reworking a program's theory to reflect changes in conditions effected by the implementation of a program.

Although the early groundwork addressing program theory incorporated the contexts into which the programs were introduced, context was factored out into the 'too-hard basket' as program evaluation worked to develop itself as a truly scientific field of inquiry. In classical experimental terms, context represents a bundle of extraneous variables to be controlled.

Pawson and Tilley (1996) point out, and a review of relevant literature confirms, that the failure of program evaluation research to address social conditions into which a program is introduced, and which continue their influence throughout the implementation of a program, is a great oversight of the field. All social programs are introduced into social contexts. If the conditions of the context into which a program is introduced are congruent with the program, the probability that the program's desired outcomes will come to fruition is greater. If the social conditions of the context into which a program is introduced are not conducive to its intended implementation, the likelihood that expected outcomes will be achieved is reduced.

Context is central to the operation of any program (Pawson and Tilley, 1996). As Pawson and Tilley (1996) point out, it is "the contextual conditioning of causal mechanisms which turns (or fails to turn) causal potential into a causal outcome" (p. 69). For example, borrowing from Pawson and Tilley (1996), when the trigger of a loaded pistol is pulled, the striking of the hammer on the bullet causes the gunpowder in the bullet to explode, launching the bullet. However, for this sequence of events to occur, conditions must be conducive. If, for instance, the gunpowder was exposed to excessive moisture at the time that the trigger was pulled, the moisture might inhibit the firing of the gunpowder.

In Trochim's (1985) opinion, the importance of context-relevant program theory warrants the incorporation of more input from program stakeholders

knowledgeable of the context or 'local' theory from which a program operates. This stand, once again, introduces the stakeholders' theory versus social science theory debate regarding the most useful source of program theory. Program theory relying on social science theory may be less specific in terms of the distinct contextual conditions of a program, yet the findings of an evaluation undertaken on the basis of such a theory should have wide application. Alternatively, a program theory generated primarily from stakeholders' views would tend to be more explicit in terms of a program's context and conditions, and evaluation findings incorporating this contextual data would be more context bound. Cook, Leviton and Shadish (1985) in their discussion regarding tactics for increasing the leverage of evaluation information, suggest grouping the individual projects of a program into classes and presenting evaluation information in those terms. The grouping of projects into classes increases the influence of the evaluation information as inferences with regards to classes rest on the demonstration of stable causal relationships across a number of projects, each with distinct characteristics (contexts). Findings supporting the success of a causative theory across a variety of sites, each with site-specific characteristics, have increased generalisability. Furthermore, evaluation information based on classes has the potential of increased leverage, as such information affects more people than information from an evaluation of a single project.

#### THE ROLE OF THE EVALUATOR

There is no consensus on how a theory-based evaluation should be undertaken (Reynolds, 1998). However, there are concerns that key stakeholders' lack of research knowledge will leave them at a disadvantage and subject to the pervasiveness and integrity of the evaluators (Patton, 1989; Stufflebeam 2001). Chen (1990b, 2004) perceives the theory-based evaluator as one among many interested parties with a stake in producing a useful evaluation of high quality. Unless there already exists a clearly explicated program theory upon which to base an evaluation, there is significant potential for an evaluator to have major influence in guiding the development of a program's causal theory (Lipsey and Pollard, 1989; Weiss, 1997a).

The evaluator has ample opportunity to influence the causal program theory upon which an evaluation is based and therefore this potential for influence is an important factor to be aware of. To consider the evaluator to be unfailingly neutral in each task he or she performs in pulling together a program's theory is fanciful. "No one prophet, intellectual or evaluator can claim to be in possession of the universal standpoint, that secret scientific key to the truth" (Pawson and Tilley, 1996, p. xii). Their values, academic training, experience, scholarly interests and professional and social affiliations almost certainly influence the nature of a theory-based evaluation (Weiss, 1997a). Furthermore, as recognised by English (2002), when evaluators generate a favourable report, their efforts are more likely to be viewed positively than if the report is unfavourable and it is unrealistic to believe that this perception will not affect the work of the evaluator. In fact, a program's theory frequently may be derived from the evaluator's own theories and experiences (Chen, 1990b; Patton, 1996) "if necessary" (Reynolds, 1998, p. 204). Pawson and Tilley (1996) caution that evaluators need to be wary of considering only information from evaluations and overlooking other rich sources of information. However, any academic interests of the evaluator should remain secondary to the stakeholders' theories (Chen, 1990b; Patton, 1996).

The conceptualisation and operationalisation skills required of evaluators conducting a theory-based evaluation necessitate that they be broadly educated and informed (Fitz-Gibbon and Morris, 1996). They need to be knowledgeable about the substantive area of the program and the program's theory in addition to knowing the design, measurement and data analysis components required for the evaluation of the program (Bickman, 1989). Broadly trained and perceptive evaluators can deduce a program's theory through observation and logical reasoning (Weiss, 1997a). Alternatively, if evaluators are working to make explicit a program's causal theory through interaction with stakeholders, they must be aware that encouraging the articulation of implicit theories may make stakeholders uncomfortable and must be particularly aware of the stakeholders' stance, while encouraging the explication of their assumptions (Weiss, 1997a). Constructing meaning through negotiation, as opposed to imposing meaning (Fay, 1977), is a skill. In short, concerns have been

expressed that the success of a theory-based evaluation rests on the evaluator's possession (at a high level) of a number and variety of skills (Bickman, 1989). Chen and Rossi (1981) appear to rely almost casually on the expertise and wisdom of an evaluator in the elaboration of a program's causal theory.

The potential of the evaluator to influence a theory-based evaluation is compounded when limited resources and time restrict the number of variables that may be studied and evaluators must realistically select a limited number of variables to be studied (Fitz-Gibbon and Morris, 1996; Weiss, 1997a). The decision about which variables to study is an "essentially subjective commitment" (Stake, 1967, p. 536). However, in the opinion of Fitz-Gibbon and Morris (1996), it may generally be assumed that the variables selected for study are those which account for the greatest variance in the outcomes of interest, and the variables dictated by the program theory to be crucial to the desired program outcomes (Fitz-Gibbon and Morris, 1996).

#### EXPLORATORY RESEARCH

In the event that no clearly explicated program theory exists, it may be decided in discussions between the program stakeholders and the evaluator that some exploratory theory-oriented research be undertaken. This inductive approach to explicating a program's theory might be undertaken during program planning or monitoring, an evaluability assessment or formative evaluation. It is a qualitative inquiry resting upon observations of the program by the evaluator (Lipsey and Pollard, 1989) for the purpose of clearly identifying, in terms of relevant contextual factors (Patton, 1996), the relationship between the program's activities and outcomes (Lipsey and Pollard, 1989). Exploratory research plays an important role in the construction of a program's causal theory when an evaluability assessment is undertaken. According to Wholey (1987), pieces of information useful in theory development are gathered when relevant documents are reviewed, during site visits, and in a series of short interviews with key program stakeholders. Continued short interviews with stakeholders further refine the causal theory (Wholey, 1987). The evaluator's organisation of the qualitative information gathered into a program theory

linking program resources, activities, intended outcomes, assumed causal links and unintended side effects is a critical step (Lipsey and Pollard, 1989).

## **IMPLEMENTATION THEORY**

A useful delineation presented by Weiss (1997b) illuminates the distinction between causative program theory and implementation theory. The definition of program theory adopted by Weiss focuses on the intervening mechanisms between intervention or program service and the intended outcomes. It is equivalent to a program's causative theory, as defined previously. Alternatively, implementation theory tests the premise that if a program is carried out as designed, the expected results will be elicited. Weiss (1997b) has termed program theories incorporating both theory types 'theories of change'. Similarly, Lipsey (1997) defines program theory as the postulations about the "change process" (p. 8) through which programs effect their desired outcomes.

From a review of published literature Weiss (1997b) finds that many evaluations stated to be causal program theory-based are actually implementation theory based. Implementation theory specifies how the program is put into action. If the program is conducted as planned and the desired outcomes occur, the implementation theory is supported. Alternatively, program causative theory focuses on the mechanisms that mediate between program implementation and the occurrence of expected results. The responses of program participants brought about by exposure to the program activities are the main focus of program causative theory, not the program activities. As Weiss (1997b) writes, "The mechanism of change is not the program activities *per se*, but the response that the activities generate" (p. 46).

Implementation theory-based evaluations commonly consider a theory of how the program operates from an observer's point of view. This implementation theory is useful for linking together the moderating variables and activities of the program in a logical manner. Stake (1967) considered that educational

evaluations conducted at that time placed too much weight on input-output relationships. He proposed an alternative where evaluation generalisations are conditional, resting upon the relationship between an educational program's contexts, activities, accomplishments and side effects which are the program's implementation theory. Certainly other evaluative researchers developed evaluative frameworks focusing on the elaboration of program implementation theory (e.g. Guba and Lincoln, 1981; Parlett and Hamilton, 1977) in an effort to illuminate complex organisational, teaching and learning processes to enable a fuller understanding of a program's operation and issues. The case study approach advocated by numerous evaluative researchers is one method (House, 1980). Other examples of early work considering moderating variables which may influence a program with regards to the program's implementation theory include Cronbach and Snow's (1977) recognition that the differences in individual aptitude may be used to predict learning outcomes and responses to variations in instructions, and Wholey's (1979) evaluability assessment work which advocates the analysis of the logical reasoning that connects the inputs of a program to the desired outcomes for the purpose of determining the compatibility of program logic and program goals prior to undertaking an evaluation. In fact, numerous evaluative researchers, including Funnell (1997) and McGraw (McGraw et al., 1996) have undertaken considerable work in the development of program logic models. These are useful elaborations of program implementation theory for the purpose of program design and evaluation that does not necessarily address the mechanisms of the program, i.e. its causative theory, but represents the logic of the program incorporating the influential moderating variables.

#### THE USE OF EVALUATION INFORMATION

A quality program evaluation intent on alleviating a difficult social problem is a task undertaken primarily to inform, and possibly improve, the social intervention under consideration. If the information from such an evaluation is ignored, the time, energy and vision dedicated to the evaluative effort, it has been claimed, have been squandered (Shadish, Cook, and Leviton, 1991a). For Patton and for many other evaluators this raises the question, "What has to be

done to get results that are appropriately and meaningfully used?” (Patton, 1997, p. 10).

In 1966, Weiss brought to the attention of her audience that underuse of findings was the most common complaint of evaluators of that time (Weiss, 1972). In response to rising concern over the apparent lack of evaluation information utilisation, a decade later, the Joint Committee on Standards for Educational Evaluation undertook a five year project that resulted in the publication detailing standards for evaluations in 1981. The set of evaluation standards called for evaluations to have four features: utility, feasibility, propriety and accuracy (Joint Committee on Standards for Educational Evaluation, 1981, 1994). The standard of ‘utility’ was purposefully listed first by the Joint Committee. This was due to their belief that an “evaluation should not be done at all if there is no prospect for its being useful to some audience” (Stufflebeam, 1980, p. 90).

Over the past four decades numerous authors have contributed to a wealth of literature on the subject (e.g. Alkin et al., 1974; Cummings, 1997; Guba, 1969; Cousins and Earl, 1992; Cousins and Leithwood, 1986, 1993; Greene, 1988; Kirkhart 2000; Leviton and Hughes, 1981; Johnson, 1998; Shadish, Cook and Leviton, 1991a; Shea et al., 1995; Shulha and Cousins, 1997; Vlahov, 1989) and to steady developments in the understanding of the use of program evaluation information (Johnson 1998). During that time utilisation has not been seriously challenged as the *sine qua non* of evaluation.

## **DEFINING PROGRAM EVALUATION INFORMATION UTILISATION**

### **INSTRUMENTAL USE**

In earlier years, Weiss’s (1972a) declaration that the provision of information for program decisions was the fundamental rationale of program evaluation, was a stand many authors of the time agreed with (e.g. Caro, 1971; Scriven, 1967). The tacit assumption was that program practitioners and policy-makers confronted with evaluation findings indicating that an intervention was either

wholly or partially ineffective, would act in such a way as to improve the program or policy that was the focus of the evaluation. This seemingly logical conception of evaluation information use is further upheld when one considers that substantial amounts of funds are directed to the conduct of program evaluations and justified by society with the expectation of some immediate return for the investment (Shadish, Cook and Leviton, 1991a). Nonetheless, early concepts of evaluation information use, later termed instrumental use, were based on a narrow definition of utilisation as the direct, immediate impact of evaluation information on discrete decisions about a program (Alkin et al., 1974). It is this narrow conception which is believed by many to have contributed to the development of the evaluation 'utilisation problem' (Cousins and Leithwood, 1993; Leviton and Hughes, 1981; Patton, 1997). Being forced to confront this 'problem', however, has led to more detailed analysis and the discrimination of various types of use.

#### CONCEPTUAL USE

Rich (1977) found that tracing a specific decision back to a particular source of information was a difficult task. However, this exercise led him to divide the use of evaluation information into two categories, instrumental use and conceptual use. Instrumental use is the direct, easily recognisable use of evaluation information to develop or improve a program, a notion which originally was recognised by early program evaluation authors as the only kind of use (Weiss, 1972b; Caro, 1971; Scriven, 1967). Conceptual use of evaluation information refers to its influence on a decision-maker's thinking about a program. Examples of conceptual use include becoming aware of evaluation information, developing an altered understanding of a program, and changes to the way a program is thought about (Johnson, 1998). Weiss (1981) endorses the idea of conceptual use, or 'enlightenment' as she terms it. Furthermore, she suggests that experiences with evaluation, and thinking about past evaluations, may accrue over time and have an influence on current decision-making, a process she calls 'decision accretion'.



## SYMBOLIC USE

By the mid 1980s further delineation of conceptual use led to the identification of symbolic use and processing use (Shulha and Cousins, 1997). *Symbolic* use occurs when an evaluation is used by individuals for political self-interest to persuade others to either support or defend a political position (Johnson, 1998). Leviton and Hughes (1981) chose to label this type of use “persuasive use” (p. 528) and point out that the interpersonal aspect of others being co-opted into accepting the implications of the evaluation information is a significant point of difference between this type of use and other uses of evaluation information. Cummings (1997) refers to this type of use as “strategic” (p. 19), a particularly appropriate term.

## INFLUENCE

Kirkhart (2000) employs the term *influence* (p. 5) to represent her expanded view of evaluation use. She suggests an integrated theory of evaluation use encompassing the three dimensions of source, intention and time to broaden the concept to include contextual considerations. In her reconceptualisation of evaluation use, the source of influence is the starting point of the change process, or the agent for change. The extent to which influence is in a particular direction, or of a particular kind, is the intention of the influence. Finally, the *time* dimension is considered in terms of immediate, end-of-cycle and long term influence recognising the concurrent, short-term and more distal effects of evaluation influence. Similar to Kirkhart (2000), other authors have expanded their conceptualisation of evaluation use to focus on evaluation influence in terms of more authentic ways of thinking about, expressing, and furthering the evidence base about the effects of evaluation and the links between evaluation and social betterment (Alkin and Taut, 2003; Henry, 2003; Mark and Henry, 2004).

## PROCESS USE

Process use, initially conceptualised by Greene (1988) in a framework for use linked to the evaluation process, is a fourth type of evaluation utilisation. Several studies have investigated the link between stakeholder participation in the evaluation process and the effect of the experience on their practice (Cousins, 1995; Cousins and Earl, 1995; Fetterman, 1994; Preskill, 1994). Patton (2004) refers to his narrow conceptualisations of evaluation use in his earlier work, claiming that this narrow conceptualisation of use resulted in the disregard of the influence of the evaluation process on the people and organisations involved. Subsequently, he has defined process use as changes in the thinking and/or behaviour of those involved in an evaluation which results from the learning that takes place in the evaluation process, and changes to program or organisational culture and processes (Patton 1997, 1998). The participation of stakeholders in undertaking an evaluation contributes to their ability to understand the evaluative process, encourages them to be more aware of inquiry practice, and enables them to consider their program and practice from an alternative perspective. From the organisational perspective, undertaking a program evaluation sends a message that the program is worth investigating, thereby potentially increasing organisational communication (Shulha and Cousins, 1997).

## **FACTORS INFLUENCING THE USE OF PROGRAM EVALUATION INFORMATION**

The concern of many evaluators with what they consider to be the under-utilisation of evaluation information became a catalyst for a search for factors that might influence evaluation information use. Weiss (1972b) aided this process by presenting a variety of factors she suggested might account for under-utilisation, in terms of instrumental use. She and other evaluative researchers were looking for predictors of evaluation information use. They thought that if evaluators were cognisant of the factors that shape the potential of an evaluation's impact, they could then incorporate them into the design and implementation of an evaluation study (Seigel and Tuckel, 1985).

The delineation of the different types of use broadened the variety of factors considered to influence the use of the evaluation information and allowed their differentiation in terms of the type of use they might affect. Later authors published alternative categorisations of factors suspected of affecting evaluation information utilisation (Alkin et al., 1985, Davis and Salasin, 1975; Patton et al., 1977). There have been a number of reviews of the research focusing on the factors which might be considered to influence evaluation information use, including those of Leviton and Hughes (1981), Cousins and Leithwood (1986), Shadish, Cook and Leviton (1991a), King and Thompson (1983), and Alkin et al. (1985). Shulha and Cousins (1997) have suggested that by 1986 numerous authors had developed categories of factors thought to influence the use of evaluation information. Three of the works presenting categorisations of such factors are considered more fully in this study. These are Cousins and Leithwood (1986,1993), Leviton and Hughes (1981) and Alkin et al. (1985).

Cousins and Leithwood (1986) reported on a meta-analysis of 65 empirical studies of evaluation information utilisation undertaken previous to 1985, from which they identified two categories of six factors as influential in their information utilisation conceptual framework. The first category of factors is concerned with the implementation of an evaluation study. The second category focuses upon the characteristics of the decision-setting. Later Cousins and Leithwood (1993) further developed their earlier framework to include a third category of factors which they identified as 'Interactive Processes'. These represent the interaction between the decision or improvement setting and the evaluation study. The factors included in each of their categories are given in Table 2.1

**Table 2.1: Three Typologies of Factors found to Influence the Use of Evaluation Information**

Cousins and Leithwood (1986, 1993)	Alkin et al. (1985)	Leviton and Hughes (1981)
<b>Decision or Policy Setting</b> Information needs Decision characteristics Political climate Competing information Personal characteristics Commitment and/or receptivity to evaluation  <b>Evaluation Implementation</b> Evaluation quality Credibility Relevance Communication quality Findings Timeliness  <b>Interactive Processes</b> Involvement Social processing Ongoing contact Engagement Diffusion	<b>Human Factors</b> Evaluator characteristics Commitment to use Willingness to involve Choice of role Rapport with users Political sensitivity Credibility Background and identity User characteristics Interest in the evaluation Commitment to use Professional style Information processing  <b>Context Factors</b> Pre-existing evaluation bounds Written requirement Other contractual obligations Fiscal constraints Organisational features Inter-organisational External features Project characteristics Age/maturity of project Innovativeness Overlap with other programs  <b>Evaluation Factors</b> Evaluation procedures Methods used Dealing with mandated tasks Use of a general model Information dialogue Amount and quality of interaction between evaluator and user Substance of evaluation information Information relevance Information specificity Evaluation reporting Frequency of information provided Timing of information Style of oral presentations Format of reports Mix of statistical/narrative data	<b>Relevance</b> Needs of clients Timeliness of reporting  <b>Communication</b> Two-way formal and informal interaction throughout the study  <b>Information Processing</b> Continuity of information flow Clarity of formal report Multiple presentation styles Qualitative and quantitative information  <b>Credibility</b> Study methodology Information provided Evaluator(s)  <b>User involvement and advocacy</b> Interest and commitment of users Turnover of users Level of user support for findings

Note: This table has been taken from Cummings (1997).

Leviton and Hughes (1981) identified 13 factors as affecting the use of evaluation information which they clustered into five major categories: relevance, communication, information processing, credibility, and user involvement and advocacy (Table 2.1). Empirical studies reported on subsequently have identified influences on evaluation use not accounted for in their categorisation of influential factors. Alkin et al. (1985) extended the number of factors identified as influential to 38 which they put into three categories: human factors, determined by characteristics of both the evaluator and the decision-makers or stakeholders; context factors, including characteristics of evaluation contractual arrangements, organisational relationships and the program; and evaluation factors including characteristics of the design, the evaluator-stakeholder interaction and evaluation reporting strategies (Table 2.1).

## **THE USE OF THEORY-BASED EVALUATION INFORMATION**

Theory-based evaluation can be viewed as yet another step in the endeavour to increase the utility and use of evaluation information. The most common reason for conducting an evaluation study using a program's underlying theory, rationale, or philosophy is to better understand the impact of the program through increased explanatory power (Smith, 1994; Weiss, 1997a). The ability of theory-based evaluations to expand intelligence as to 'why' a program works, in addition to providing information regarding 'whether', or 'to what extent' a program works, typically produced by traditional 'black-box' evaluations, is claimed to lead to an increase in the usefulness of program evaluation information in terms of program and policy development (Chen, 1990b, 2004; Pawson and Tilley, 1996). Therefore, it is interesting that a review of the evaluation information utilisation models published to date has not identified a single one that has considered the influence of theory-based evaluation practice on evaluation utilisation. This dissertation was undertaken in order to do so.

The investigation of the influence of theory-based evaluation on the use of evaluation information in terms of most established utilisation models, does

not require a redevelopment of the model. Theory-based evaluation is an alternative form of evaluation practice that produces information reflecting the program's theory, including the program's prescriptive, causative, and implementation sub-theories. The literature suggests that the information content of theory-based evaluation studies is different, in terms of theory orientation, to other traditional evaluation practices. Following on, relevant literature indicates theory-based evaluation is more useful to stakeholders because of the theory orientation (for example, Bickman 1987a, 1987b; Chen, 1990b; Pawson and Tilley, 1997). However, the types of use identified in the evaluation utilisation literature remain unchanged. Furthermore, in the current evaluation utilisation model, the evaluation categories and the characteristics of factors commonly recognised in published literature as influential on the use of evaluation information remain unchanged, even with regards to theory-based evaluation. However, because this is an application of the evaluation utilisation model to an alternative practice, some pertinent points for consideration are presented below.

The theoretical structure of a program dictates the nature of appropriate measures to be included in a theory-based evaluation, guiding the collection of well-focused information. Without a theory to provide guidelines it is difficult to determine which variables would provide relevant measures of a program's outcomes and which would not (Bickman, 1987). Conrad and Miller (1987) are of the opinion that the development of measures of a program's causal links requires a setting of standards that may initiate program improvement prior to the actual undertaking of a theory-based evaluation. A program theory can provide information regarding sub-goals that must be accomplished before the intended outcomes may be successfully achieved. For example, a program intending to change people's eating habits to include healthier food choices might propose that the target population first be educated to recognise healthy foods from unhealthy foods. A formative evaluation might detect that a significant proportion of the program participants are not educated in the choice of healthy foods, thus providing information to bring the program, as implemented, into line with the program theory (Bickman, 1987). Furthermore, evaluating a program in accordance with an underlying theory

enables the evaluator to look beyond the intended program outcomes specified by program developers to effects that can be inferred from the theory in operation.

A program's causative theory identifies the causal links between the program application, or intervening variables, and its intended effects. Identifying these intervening variables and carefully considering their effect in the implementation of the program can provide policy-makers and program developers with useful information regarding the program operation. Implementation process theory (Scheirer, 1987), alternatively, addresses the variables affecting the extent to which the program is implemented as planned. The delineation of program theory failure from program implementation failure, is not possible unless it can be shown that a program was faithfully implemented in accordance with the implementation theory underpinning it.

Theory-based evaluations highlight the relationships between the identified need, the program operation, the target group and the intended outcomes of the program. A program is developed with a particular identified need and a target population in mind. It may be the case, however, that the program is applied to an inappropriate problem or target group. An enhanced likelihood of discerning between program theory failure and implementation theory failure, is a significant merit of theory-based evaluation addressed by numerous authors (Bickman, 1987; Scheirer, 1987; Weiss, 1997a).

If a theory-based evaluation finds that a successful program has been implemented in accordance with the program theory, then the evaluation information lends support to the causative theory. However, should the theory-based evaluation find that the implementation of the program is not as intended, success or failure could be a result of the way the program was actually implemented, rather than the causative theory as espoused. As Bickman (1987) has revealed, there are numerous examples in the literature of interventions or programs found to have no effect and hence discredited, when in fact the intervention was never implemented as intended and never tested.

Many authors have written of micro and macro program theories recognising the basic dissimilarities in perspective and priority between the micro and macro program levels (Cook, Leviton and Shadish, 1985; Scheirer, 1987; Shadish, 1987). For example, an evaluator approaching a program at the national level will most probably be adopting a macro level program theory, focusing on gathering implementation process information as the program proceeds down through bureaucratic levels to implementation at the micro level. Programs conceptualised at the national level typically are applied across a variety of sites, times and participants. Put another way, programs provide the administrative link between policies conceptualised at the national level and services delivered at the local level, through the appropriation of funds (Cook, Leviton and Shadish, 1985). Scheirer (1987) believes auditing type measurements are a favoured characteristic of this level of theory-based evaluation.

In contrast, micro level theory-based evaluations incorporating characteristics of the structure and operation of a program, including activities and immediate program outcomes, are undertaken from a local perspective and focus on program elements. Program elements are the components required for the achievement of a program's intended goals (Cook, Leviton and Shadish, 1985). For example, an element of a program intended to change people's eating habits would be the education process to promote healthy foods.

An evaluation may address program theory at both the macro and micro levels, presenting a complete model of a program's causative theory and implementation theory, and meeting a range of stakeholders' information needs. However, the reality is that in spite of the interrelation of program levels, when an evaluation is undertaken one level usually is chosen as the focus of the evaluation. The selection of the evaluation level has important implications in terms of information needs.

The increased generalisability of evaluation information based on program theories incorporating social science theory is particularly useful for policy-makers and academics. Understanding the constructs underpinning a program



should allow some freedom to predict the success or failure of programs implemented differently in terms of population, times or location, thus furthering the theory at hand by testing its external validity (Bickman, 1989; Chen and Rossi, 1981). The inference that different program operational definitions would produce similar effects is another sort of generalisation known as ‘construct of cause’ validity, in which the independent variable, the intervention or program, is varied, testing the robustness of the theory (Cook and Campbell, 1979).

Theory-based evaluation offers an ability to inform social science theory. However, for an evaluation of a program’s theory to make such a contribution it is imperative that the independent and the dependent variables are operationalised well and the construct validity of the measures is sound (Bickman, 1987). Lipsey (1997) envisions theory-based evaluations as building blocks for the development of better social intervention theories useful for guiding program development and evaluation designs.

There are some concerns that theory-based evaluations may have a negative effect on the use of evaluation information. Patton (1989) writes of concerns that information from evaluations conducted in accordance with a theory primarily derived from a social science theory, as initially proposed by Chen and Rossi (1987), may be too esoteric and abstract for stakeholders to make much use of. Furthermore, he is concerned that the level of research sophistication required by such an evaluation may alienate stakeholders from the evaluation process.

Scriven (1994) is concerned that an evaluation addressing a complex theory regarding the relationships linking the program components, or the internal operations of the program, may not be accurate and queries the means by which such a program theory may be ascertained. He feels information useful to stakeholders comes from sound, logical evaluation concerned with the establishment of a program’s “merit, worth, quality, or value” (Scriven, 1994, p. 75) and feels evaluation can be very useful without ‘risky’ undertakings such as program explanation or recommendations.

## FACTORS AFFECTING THE IMPACT OF AN EVALUATION

Essentially, an evaluation study can be conceived of in terms of a large number of factors, each with an influence on the evaluation process and the outcome. Weiss (1972b) seems to have initiated the practice of categorising the factors using instrumental use as her perspective. Her intent was to provide insight as to why evaluation studies were often not used to make direct immediate decisions.

Numerous other authors have investigated factors considered by many to have a significant impact on the effects of a study (Alkin et al., 1988; Cousins and Leithwood, 1986, 1993; Leviton and Hughes, 1981). The factors delineated in the relevant literature include the methods of the study, the study process, the quality, relevance and timeliness of the evaluation information to the program needs, the quality of the evaluator-stakeholder communication, the format of the report, and the context of the study including the political environment. Essentially, most studies suggest that the absence of an identified factor or its poor quality increases the likelihood that the evaluation will have reduced or no impact in terms of the use of its information.

As the results of the empirical studies regarding identified factors increased so did the number of categories. Leviton and Hughes (1981) identified five categories for grouping the factors, i.e. relevance, communication, information processing, credibility, and user involvement and advocacy. Alkin et al. (1985) identified 38 factors which they grouped into just three categories: human factors, context factors and evaluation factors, while Cousins and Leithwood (1986), in their review of empirical studies of evaluation utilisation, identified only two categories, evaluation implementation factors and decision-or policy-setting factors. Cousins and Leithwood (1993) later added a third category, recognising the interactive processes between those responsible for disseminating the evaluation information (e.g. evaluators) and those responsible for using the information (e.g. stakeholders).

Although the factors are grouped differently into their categories they can generally be divided into common areas of focus. These are characteristics of the evaluation stakeholder group, the evaluator, the evaluation process and the environment of the evaluation, i.e. its context. These areas are considered below with specific reference to theory-based evaluations.

### **Stakeholder Characteristics**

Patton et al. (1977) identified what they called the ‘personal factor’, “...made up of equal parts of leadership, interest, enthusiasm, determination, commitment, aggressiveness and caring.” (p. 73) of individual people as being of importance to utilisation. They found that when the ‘personal factor’ was evident, evaluation had an impact; when it did not, impact was generally absent. With regards to the impact of evaluations in general, the findings of Patton et al. (1977) suggest the positive influence of stakeholder commitment to an evaluation and to the use of the evaluation information.

A positive relationship between stakeholder commitment to the evaluation, defined as involvement, and evaluation information use, has been identified by numerous authors (Alkin et al., 1979; Brett, Hill-Mead and Wu, 2000; Cousins and Earl, 1992, 1995; Cousins and Leithwood, 1986; Cummings, 1997; Dawson and D’Amico, 1985; Dickey, 1980; Greene, 1988; Hudson-Mabbs, 1993; Patton et al., 1977; Patton, 1997; Preskill and Caracelli, 1997; Rich, 1979; Vlahov, 1989). Communication between those responsible for the evaluation and the evaluation stakeholders as a particular tool of involvement, has been, and continues to be, considered an area influencing utilisation (Cummings, 1997; Greene, 1988, 2002). A number of reasons for the strong link between stakeholder involvement in the evaluation and the use of its information have been identified. It has been suggested that the involvement of stakeholders in the evaluation process increases the likelihood that the information needs of the stakeholders will be met, making the evaluation information more relevant to them (Cummings, 1997; Patton, 1997). Relatedly, Cousins and Walker (2000) found teachers’ sense of personal efficacy to be a variable predictive of attitudes towards systematic inquiry,

supporting the findings of previous studies which indicated that high efficacy teachers are more receptive to new innovative, challenging, change-oriented, collaborative and academically-oriented practices and teaching techniques (Guskey, 1988; Smylie, 1988).

It has been suggested that the involvement of stakeholders in the evaluation process prepares them for accepting the information by making them aware of the program context and the evaluation processes. This involvement, it is proposed, increases a stakeholder's understanding of the evaluation findings, ownership of the evaluation information, sense of personal responsibility for advocating the evaluation information, and likelihood of them accepting the information as valid and credible (Cummings, 1997; Greene, 1988; Preskill and Torres, 1999). It has also been proposed that the involvement of stakeholders in the evaluation process helps to create interpersonal networks necessary for the sharing of information, meaningful discussion and reflection (Cousins and Earl, 1995; Louis and Simsek, 1991). Furthermore, strategies to increase stakeholder involvement in evaluation processes likely to increase stakeholder use of the study information have also been linked to the development of organisational learning (Cousins, 1996; Leithwood and Louis, 1999).

Some studies indicate previous education and experience regarding research, positively influences stakeholder opinions of it, their perceptions regarding their understanding of it and their ability to undertake it (Cousins and Walker, 2000; Greene and Kvidahl, 1990). Furthermore, previous participation in research was found by Cousins and Walker (2000) to be a significant predictor of attitudes towards research. They found that when educators had undertaken prior research-related coursework they had more favourable attitudes towards research.

Stakeholders' perspectives directly influence the likelihood of them using information derived in an evaluation based on a program's theory (Patton, 1989). As Cummings et al. (2001) write in their paper on the use of program theory in an educational setting, the final description of the program serves to

inform stakeholders about the program evaluated and is critical to both the quality of the evaluation study and the manner and degree of the evaluation study information influence. Patton (1989), however, is concerned that stakeholders with a limited research background may be intimidated by the conceptualisations and terminology adopted in a theory-based evaluation, thereby decreasing the likelihood that the evaluation information will be used. Furthermore, stakeholder expectations have an influence on the impact of the evaluation information. The greater the extent to which an evaluation meets stakeholder needs and expectations of information regarding a program, the greater the likelihood that the evaluation's information will have an impact (Alkin, 2004; Chen 2004).

In undertaking a process to make overt the program theory held implicit by stakeholders, the opportunity occurs for “social constructivist learning” (Preskill and Torres, 2000, p. 31) where individuals consider information together and have an opportunity to reach some consensus regarding the program theory adopted for the evaluation, as well as to consider concurrently the information needs of the program. The literature suggests that the involvement of stakeholders in the theory elaboration process and the process of achieving consensus regarding the program theory, may increase their awareness and understanding of the program, encourage reflection regarding their practice, lead to changes in their practice, and inform decisions regarding the program evaluated (Cummings et al., 2001; Huebner, 2000; Milne, 1993). In fact, Fetterman (2004) in his elaboration of ‘empowerment evaluation’ suggests that enabling stakeholders to pursue a sound understanding of the program, the environment and the theory of action of the program (program theory as espoused) and theory of use of the program (program theory in reality) is an initial step towards empowerment.

### **Evaluator Characteristics**

An evaluation, like any other social process, is a product of the people involved, the program practitioners, the clients, the stakeholders and the evaluators. No two people will perform their role in the same way. In 1976

Meltsner, based on interviews with 116 federal policy analysts in the United States, concluded that evaluators differ in expectations, training, norms and motivations.

Evaluators are responsible for undertaking evaluations with a potential for powerful influence. Numerous authors have considered the influence of the characteristics of the evaluator, or evaluation team, on the evaluation process and outcomes (Braskamp et al., 1982; Conley-Tyler, 2005; Cummings et al., 1988; Greene, 1988; House and Howe, 1998; Lake, 2005; Mathison, 1994; Owen, 2006; Scriven, 1991; Weiss, 1972b). The belief that the interest, ideology and background of an evaluator have an impact on the use of evaluation information is supported by numerous authors (Alkin and Dailak, 1985; Cousins and Leithwood, 1986; Greene, 1988; Hammond, 1983; Lake, 2005; Mathison, 1994; Seigel and Tuckel, 1985). In fact, Hudson-Mabbs (1993) found the credibility of the evaluator was particularly important where potential dissension exists between interested parties. In particular, the stakeholders' perception of the evaluators has been found to have an influence when there is potential for dissension among the stakeholders of the evaluation regarding the findings of an evaluation study and when the findings are to be disseminated outside the program (Hudson-Mabbs, 1993; Leviton and Hughes, 1981).

In his consideration of evaluator practice in the 'ideal evaluation' Alkin (2004) suggests themes which can be applied to any evaluation approach with a view to enhancing evaluation use, all resting on the skill and abilities of the evaluator. First, he suggests that evaluators focus on utilisation of the evaluation's information and undertake to increase stakeholder perception of potential avenues of use of the evaluation at the conceptualisation stage of the investigation. Second, he believes the evaluator's willingness and ability to be flexible in terms of making modifications to the evaluation as the need arises in the process, is an avenue to increased evaluation usefulness. Third, he advocates that evaluators concentrate on developing a significant, positive relationship with potential users of the evaluation (usually the stakeholders) in an effort to enhance 'buy-in'. He suggests all three of these practices involve a

substantial amount of work prior to the formal collection of data for the study. Irrespective of the form of the evaluation approach, theory-based or otherwise, Alkin's themes apply, highlighting the potential for influence of the evaluator's skills and abilities on the influence of the evaluation's information. As a result of this foundation, the evaluator of the 1970s often was charged with the task of conducting evaluations for the purpose of providing program stakeholders with information for decision-making. However, in the late 1970s and early 1980s evaluation theorists began to advocate more strongly for the evaluator undertaking an educator role (Cronbach et al., 1980). In this role the evaluator is seen to become responsible for providing stakeholders with an understanding of their program and the evaluation process. Then, roughly a decade later stakeholders were brought a step further into the evaluation fold and encouraged to participate in the evaluation process (Greene, 1987). With this change, the evaluator adopted the role of collaborator and facilitator in the construction of the evaluation process, creating avenues for stakeholder participation (Mathison, 1994).

As presented earlier in this chapter, a number of authors have written about the role of the evaluator in undertaking a theory-based evaluation (Bickman, 1989; Chen, 1990b; Fitz-Gibbon and Morris, 1972; Lipsey and Pollard, 1989; Patton, 1996; Pawson and Tilley, 1996; Reynolds, 1998; Weiss, 1997a). A successful theory-based evaluation rests heavily upon the evaluators' skills which increases the influence of the evaluators' characteristics on the usefulness of the evaluation information (Bickman, 1989).

As suggested by Stufflebeam (2001), in undertaking a theory-based evaluation there is an opportunity for the evaluator to develop their own program theory. In the program theory elaboration process, the evaluator also has an opportunity to have a significant influence on the program theory agreed upon by the stakeholders. Nevertheless, the evaluator should bring to the exercise a good knowledge of the relevant social science theory, the success of the process relies on the skill of the evaluator and their awareness of their own values and experience in managing the process (Weiss, 2000). Furthermore, Stufflebeam (2001) has also cautioned that evaluators might, in the evaluation

process, displace whatever program theory staff members have been using to create the program design with their own version. Chen (2004), however, counters that an evaluator's role is to assist or facilitate stakeholders in making explicit their implicit theory of the program. The success of the exercise rests on the evaluator's ability to document the program theory explicated by the stakeholders and implement it systematically.

### **Characteristics of the Evaluation Study**

Numerous characteristics of an evaluation study have been identified as having influence on the effects of an evaluation's information. These include, for instance;

- i. communication quality (Alkin et al., 1985; Burry et al., 1985, Dickey, 1980; Greene, 1988; Marsh and Glassick, 1988, Cummings, 1997),
- ii. credibility of the evaluation process (Alkin and Daillak, 1985; Seigel and Tuckel, 1985),
- iii. quality of evaluation study methodology (Alkin and Daillak, 1985; Dickman, 1981; Patton et al. 1977),
- iv. findings of the study in terms of nature of the results,
- v. consideration of program finances and resources,
- vi. congruence with stakeholder's expectations,
- vii. value for decision-making (Braskamp, Brown and Newman, 1982; Davis and Salasin, 1975; Hammond, 1983; Lorenzen and Braskamp, 1978; Seigel and Tuckel, 1985),
- viii. relevance of the information to the needs of those responsible for putting the information to use (e.g. Alkin et al., 1985; Dickman, 1981, Hammond, 1983; Weiss, 1972c),
- ix. timeliness of evaluation information (e.g. Patton et al., 1977; Seigel and Tuckel, 1985),
- x. commitment of stakeholder to the evaluation (e.g. Greene, 1988; Hammond, 1983; Vlahov 1990), and
- xi. competing information (Alkin and Daillak, 1985; Vlahov, 1990, Newman et. al., 1983).



A consideration of the program's theory in undertaking an evaluation has potential to affect all characteristics of the evaluation process. Theory-based evaluations highlight the relationships between the identified need, the program operation, the target group, and the intended outcomes of the program. A program is developed with a particular identified need and a target population in mind. It may be the case, however, that the program is applied to an inappropriate problem or target group. Discerning between program theory failure and implementation theory failure is a significant merit of theory-based evaluation addressed by numerous authors (Bickman, 1987; Scheirer, 1987; Weiss, 1997a).

The theoretical structure of a program dictates the valid measures to be included in a theory-based evaluation, guiding the collection of well-focused information. Without a theory to provide guidelines, it is difficult to determine which variables would provide valid measures of a program's outcomes and which would not (Bickman, 1987). Conrad and Miller (1987) are of the opinion that the development of measures of a program's causal links requires a setting of standards that may initiate program improvement prior to the actual undertaking of a theory-based evaluation. A program theory can provide information regarding sub-goals that must be accomplished before the intended outcomes may be successfully achieved.

### **Evaluation Context**

Pawson and Tilley (1997), in their book *Realistic Evaluation*, have advocated that in undertaking an evaluation study of a program the relationship between the causal mechanisms of the program, and their effects be considered. However, they recognise that the relationship between the causal mechanisms and their effects is neither fixed nor definite. It is contingent upon the presence of 'contextual conditions' to change the potential for a casual effect into a reality (p. 69). In the present consideration of organisational characteristics and the impact of an evaluation, the essence of the *Realistic Evaluation* concept of 'contextual conditions' applies.

Recognising the powerful influence of the organisational culture and norms on the access and use of research and evaluation information, a number of authors recognise the importance of the organisational environment as a contextual influence on evaluation use (Cousins and Walker, 2000; Cousins and Earl, 1995; Preskill and Torres, 1999a, 1999b, 2000; Jenlink, 1994; Mathison, 1994; Owen and Lambert, 1995). For use of information to occur, both conceptual and operational opportunities to attend to and apply the information must be available to the individuals, as only then is larger scale use of information likely to occur (Kim, 1993; Preskill and Torres, 1999). Preskill and Torres (1999a, 1999b, 2000a) argue that for evaluative inquiry to be successful, an organisation must have the infrastructure to support and facilitate it.

The logic of including organisational characteristics is essentially thus: if the organisation responsible for the delivery of a program, and the focus of an evaluation, has structures and practices in place which inhibit the reception and use of the evaluation information, the potential for use is diminished. Accounting for an organisation's environment, in which the evaluation's information is expected to have an influence, enables an awareness of both the independent and extraneous organisational barriers that may influence evaluation use. The implications of the logic expressed in much of the literature regarding organisational learning for evaluation practice are powerful. Simply, the greater the extent to which an organisation has a learning nature, the more likely its members are to attend to a range of information, including evaluation information, in the development and conduct of the program they deliver, and the more likely it is that they will be in a position to use that information in the decisions which they make.

The concept of organisational learning has been in the literature for thirty years (Argyris and Schön, 1978), but the concept of a learning organisation is more recent (Senge, 1990; Watkins and Marsick, 1992). Learning organisations, as popularized by Peter Senge (1990), are places "where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together" (p. 1).

Although the concept of a learning organisation has been popular for over a decade, the definitions offered by scholars and academics to capture the meaning and spirit of the 'learning organisation' have been abstract and "filled with near mystical terminology" (Gavin 1993, p. 78). In his conceptual article reviewing successful and unsuccessful implementation of learning practices in organisations, Garvin (1993) suggests organisations with a commitment to learning are good at "...systematic problem-solving, experimentation with new approaches, learning from their own experience and past history, learning from experiences and best practices of others, and transferring knowledge quickly and efficiently throughout the organisation" (p. 81). The organisational learning of interest is a continuous process integrated into organisational practices, not short-lived, spurious nor episodic.

Watkins and Marsick (1992) in their book offering strategies for the structuring of learning organisations, advocate six requirements. These requirements are: the creation of continuous learning opportunities; the promotion of inquiry and dialogue; the encouragement of collaboration and team learning; the establishment of systems to enable shared learning; the empowerment of organisation members in the development of a shared vision; and the connection of the organisation to its environment. The presence of these characteristic strategies supports the development of learning groups with a shared understanding of the path to job completion and encourage the development of commonalities in the way members talk and think about their work. In the opinion of Stamps (1997), it is in these groups that the most valuable and innovative learning takes place. Organisation members of learning communities are likely to pursue issues of common interest and practice, and seek consensus in decision-making.

## CONCLUSION

Historically, the usefulness of evaluation information has been the measure considered most important in the judgment and development of program evaluation. If evaluation information does not have an impact, it was claimed, there was little point in doing an evaluation study. The history of the evolution

of program evaluation provides the context for the emergence and elaboration of the program theory-based evaluation approach; an evaluative approach aimed at increasing the usefulness and use of the evaluation's information. Although significant conceptual literature addressing the positive influence of a theory-based approach on the impact of an evaluation may be found in a review of the relevant literature, little empirical work has been undertaken to investigate the claim.

This chapter considers the emergence of the theory-based evaluation approach and outlines the early influences on both the more traditional and other non-traditional forms of program evaluation. The conceptualisation and definition of what is meant by the effect or 'use' of an evaluation's information is then focused upon. The history of the use of program theory in program evaluation then follows, which begins with descriptions of program theory by type, source, and consideration of context. These concepts are the categories of the program theory Classification Matrix developed in the following chapter for the classification of program theory use in evaluation studies. Literature addressing the merits and criticisms of undertaking an evaluation guided by a program's theory, either causative or implementation based, is also considered. The final sections of the chapter address the use of evaluation information, first defining the various types of use identified in the literature including instrumental, conceptual, process, symbolic use, as well as the influence of evaluation information. The effect of the use of program theory is then considered in terms of the effect of the process on specific factors indicated by research to have an impact on the use of an evaluation.

In the following chapter, the factors considered to have impact on the use of an evaluation and its information is again focused on in the development of the eight versions of the Core Model of the study which are to be tested in this investigation. The factors are operationalised as predictor variables of the outcome model which is use of the evaluation information. Comparisons of model fit are considered with a view to gaining insight into the influence of the predictor variables on the outcome variations.

## **CHAPTER III**

### **MODEL DEVELOPMENT**

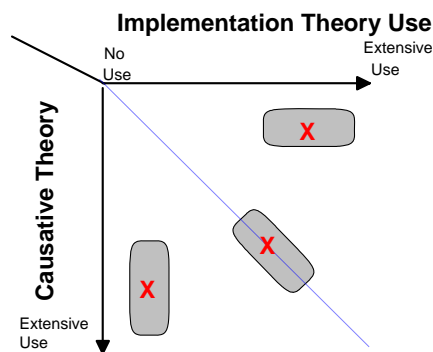
#### **INTRODUCTION**

Consensus about how to conduct a theory-based evaluation is lacking and guidelines pertaining to the conduct of theory-based evaluations are indefinite (Reynolds, 1998). The lack of specific formalized guidelines for determining what is, or is not, program theory encourages practitioners to attach their own meanings to the same words (Weiss, 1997b). Therefore, an initial step in the development of a model for investigating the influence of program theory use in an evaluation, is to understand what is meant by a program's theory in terms of how it evolves and the significant influences.

This chapter contains five sections, each focused on the elaboration and explanation of a different concept related to theory-based evaluations. The first applies Straton's (1990) four stage model of the evaluation process to theory-based evaluations. The second presents a data collection matrix, to guide the collection of data in undertaking a theory-based evaluation based on the 'Descriptive Matrix' portion of Stake's data collection matrix originally presented in his 'Countenance Paper' (Stake 1967, p. 529). The third section presents an evolution process of a model in stages, to illustrate the relevant influences on the development of a generic program's theory. This section ends with a consideration of the effect of an evaluation on the program's theory. The fourth section presents the Core Model developed for testing in this study. The eight model versions to be tested and further elaborations of the model are discussed. The final section of the chapter focuses on the description and application of the data collection matrix developed to classify evaluation studies in accordance with their use of program theory source, type and degree. Twenty one studies are described and classified using the matrix.

A primary consideration of this study is the influence of the use of theory-based evaluation practice on the use of evaluation information. The study

includes three case studies of concurrent evaluations of ongoing programs with a view to identifying the use of the information of each evaluation, and determining the characteristics of the cases which influence that use. One of the evaluation studies included, is based on the strong use of program implementation theory. To address the other end of the matrix, an evaluation based on the strong use of program causative theory undertaken without regard for program implementation theory, is considered. The final study included falls somewhere between the two extremes of theory use in evaluation practice marked by the first two. The variations in theory use among the evaluation studies included enable the comparison and contrast of the influence of theory use, by degree and type, on the use of the evaluation information. The targeted program theory areas are represented in Figure 3.1 below.



**Figure 3.1: Theory Use of Evaluations Targeted for Inclusion in Study**

## THE EVALUATION PROCESS

Straton (1977, 1990), drawing on the definition of evaluation of Stufflebeam et al. (1971) identifies four stages of the evaluation process, i.e. Delineating, Obtaining, Providing, and Utilising. He argues these stages represent the basics of any evaluation study, although he recognises that some or all of these stages may be repeated in certain types of evaluation studies.

The Delineating stage is focused on processes such as the identification and prioritisation of the evaluation study audiences, the determination of their

information needs, a consideration of the purposes of the study and deciding what information to gather.

The Obtaining stage involves getting the information and encompasses four undertakings. First, existing information is identified and methods for its collection and assessment determined. Second, the design of the evaluation study is determined with due consideration of factors such as the acceptability of the design to the evaluation audiences and clients, the cost of the approach and the likely influence of the design type on the use the evaluation information. Third, the choice of data collection methods, whether to adopt available instruments or develop one or more, consideration of the study population, and associated costs are pertinent to the instrumentation for the study. Finally, the information is obtained; available information is collected and study instruments, either adapted or developed for the particular study, are administered.

The Providing stage is concerned with data collection, documentation and dissemination methods. The undertaking of the data analysis includes the selection of appropriate statistical tests and techniques based on considerations such as the meaningfulness of the resultant statistical information to the audiences and the costs associated with the techniques. Documentation type is determined considering the audience characteristics, associated costs and production time, and the expected influence of the type of documentation on information utilisation. Dissemination mode is again a reflection of what best suits the primary audiences, associated costs and time. Other pertinent considerations include the order of dissemination to the various audiences and the influence of dissemination on the use of evaluation information.

The Utilising stage is orientated to achieving optimum use of the information and is characterized by a consideration of the likely uses of the evaluation information, both conceptually and instrumentally, factors influencing the use of the evaluation information, such as the characteristics of the decision-making/judging setting, and the characteristics of the evaluation study.

A representation of the four stages of the Evaluation Process Model, taken from Straton (1990), is presented in Figure 3.2. The model detailing the evaluation process presented in Figure 3.3 is a generic representation of the evaluation process in that the fundamental steps required in any evaluation are represented. However, evaluation studies resting upon these basic steps may adopt different forms or approaches that would build upon the basic evaluation process (Owen, 2006). The evaluation approach of particular interest in this dissertation is program Theory-based Evaluation. The following section considers the process of Theory-based Evaluation.

### **THEORY-BASED EVALUATION PROCESS MODEL**

Many evaluators have argued that using the theory of a program to guide the evaluation study provides important clues as to why a program fails or succeeds and provides evidence of whether the observed outcomes can be attributed to the program (Rogers et al., 2000). Non-theory-based evaluations may consider program inputs and outputs, but not the underlying mechanisms of the processes linking them. With interest in, and use of, the theory-based evaluation approach growing amongst evaluation practitioners, it is necessary for information regarding the value of the process to evaluation practice to be available. Before this can come about, steps must be taken to explicate the theory-based evaluation process. The intent of this section is to articulate the theory-based evaluation process.

For the purposes of this study, the Evaluation Process Model (Figure 3.2) presented by Straton (1990) has been adapted to represent a Theory-Based Evaluation Process Model (Figure 3.3). The theory-based version differs from the Evaluation Process Model with the inclusion of the Program Theory Articulation Process which influences the four essential evaluation study stages. In a theory-based evaluation process the program theory articulation process originates in the Delineating stage, and is then operationalised and implemented in the Obtaining and Providing Stages and, it is proposed, influences the Utilisation stage. A discussion of the activities of the four stages of the evaluation process is undertaken below.

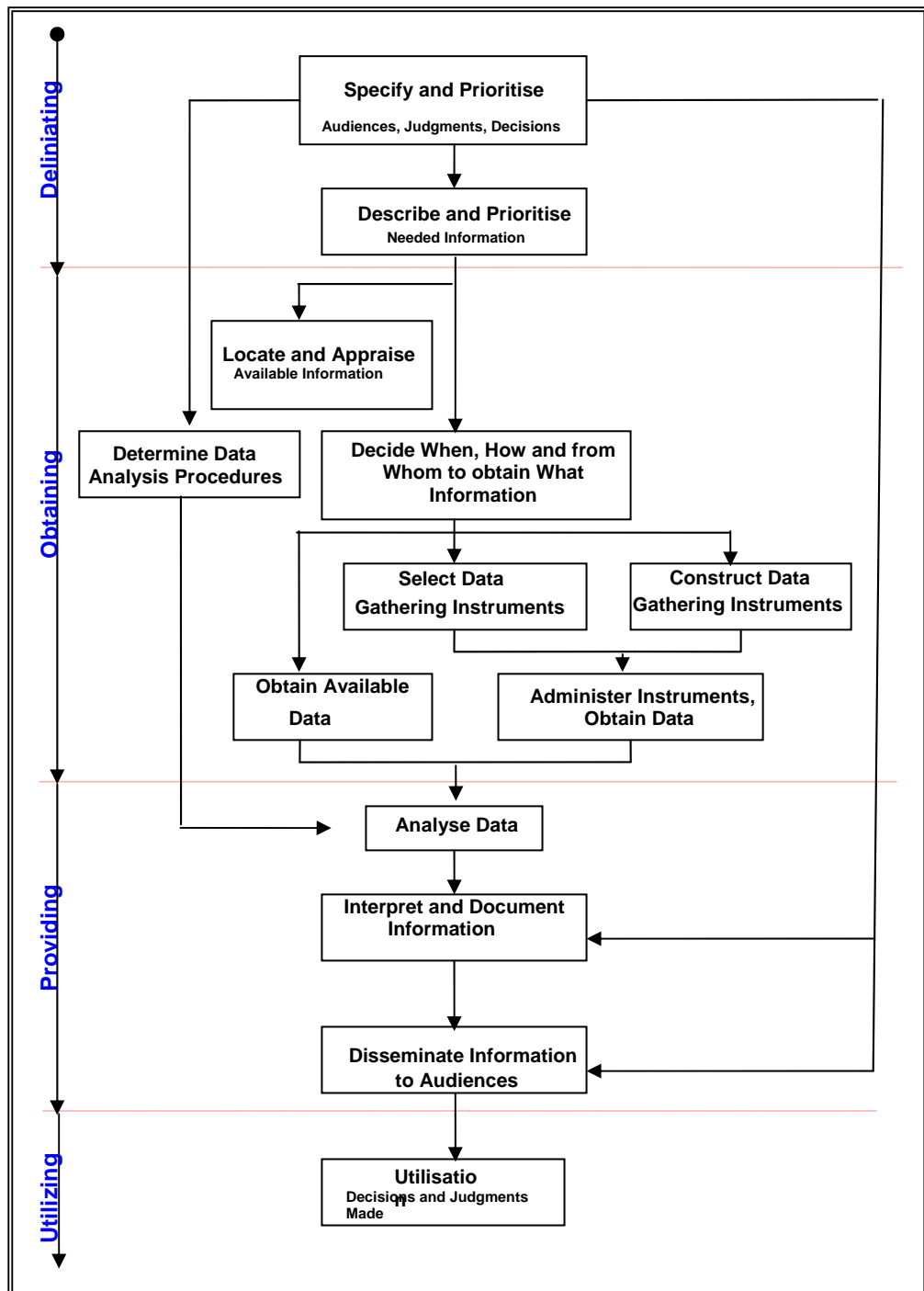


### **Specify and Prioritise**

The initial considerations in the evaluation process relates to the specification and prioritisation of the evaluation audiences, evaluation purposes and information needs. This is primarily a conceptual exercise undertaken by the evaluator, based on information from key informants, program documents and observation. Essentially, evaluation audiences are identified and their information needs considered. Often a prioritisation or ranking of pertinent audiences occurs, which is heavily influenced by who the evaluation study has been contracted by. The purposes of the evaluation, both formal and informal, are determined. Again, those responsible for the contracting of the evaluation typically have some very definite ideas regarding the purpose of the evaluation. Certainly, any pending decisions or judgments should be considered. A complete model of the Theory-Based Evaluation Process is represented in Figure 3.3.

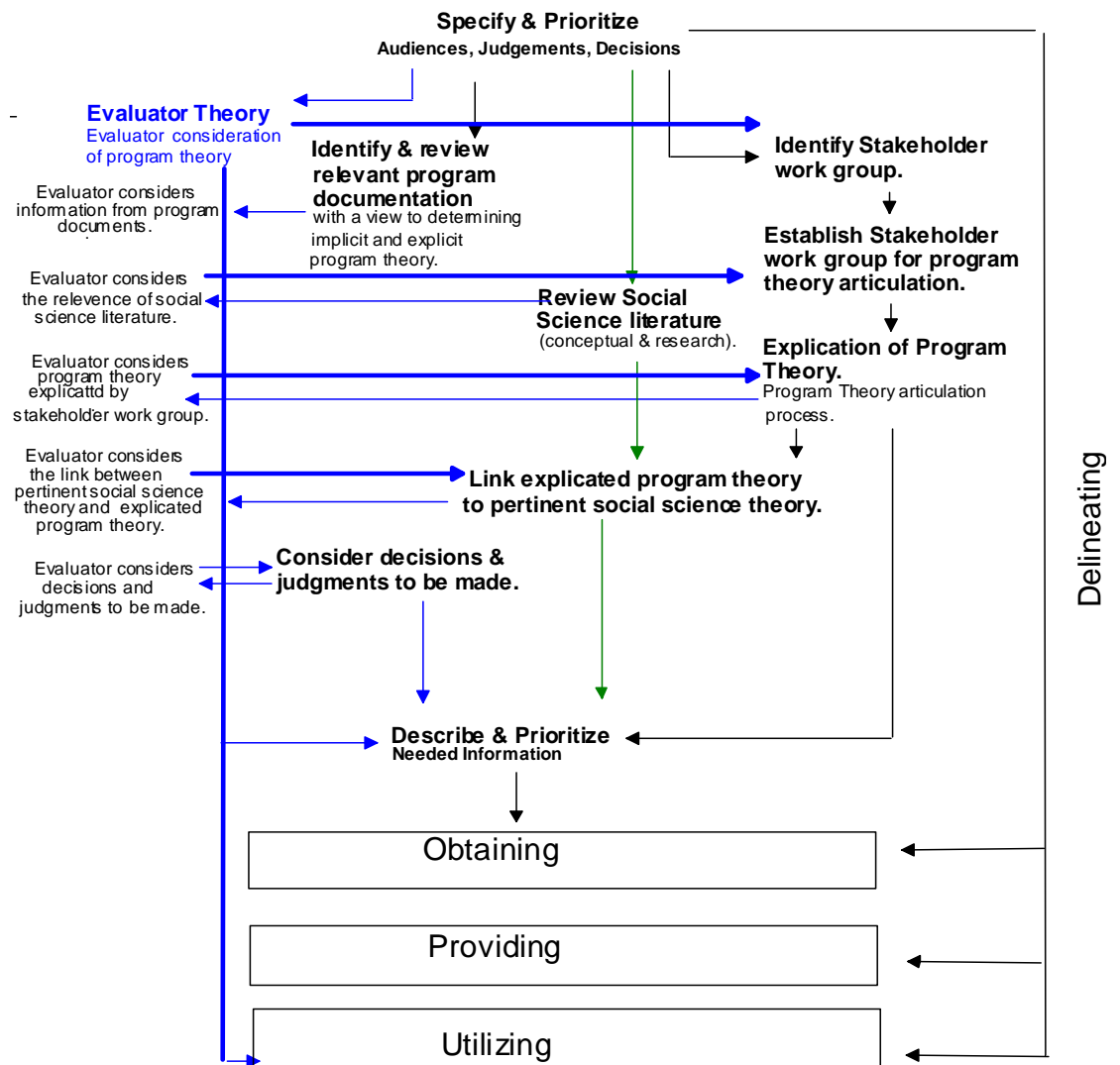
### **Program Theory Articulation**

The three primary origins of a program theory articulated for theory-based evaluations are the evaluator, social science theory and the stakeholders. The contribution made by evaluator theory and stakeholder theory to the program theory articulation process are considered below. Social science theory, although a significant source for program theory articulation, is included through evaluator and stakeholder theory articulation. It is important to note that this is only one approach to undertaking theory-based evaluation. A review of pertinent literature reveals a wide variety in the practice of theory-based evaluation studies (e.g. Bickman, 1996; Chen, Wang and Lin, 1997; Goodman et al., 1996; Mertens, 1996).



Note: This model has been taken from Straton (1990). However, the four stages named along the left-hand side of the figure have been incorporated into the model, in this instance.

**Figure 3.2: A Model of the Evaluation Process**



**Figure 3.3: Theory-Based Evaluation Process Stages**

### Evaluator Theory

The initial step in the consideration of an program evaluation is primarily a conceptual one, as noted earlier. The evaluator first looks at why the evaluation is being undertaken and then considers the nature of the evaluation which is needed. Determining the nature, purpose, and information which needs to be addressed by the evaluation initiates the evaluator's consideration of the logic of the program, may prioritise the usefulness of the information contained in certain program documents in comparison with others and guides

the literature review process. Audiences are identified and pending judgments or decisions considered. These are contextual markers. The establishment of these points sets guidelines for the investigative process. Although the program stakeholders, whose views will be reflected to some extent in the program documents, and the social science literature, are important origins of the program theory developed for the evaluation, the evaluator is a third origin of program theory whose influence is less easily traceable, yet possibly more pervasive.

The evaluator's theory represents their view of the underlying rationale of a program including the mechanisms by which it is thought that the program objectives will be attained. It is informed by a review of relevant program documents, a review of pertinent social science literature, any program theory explication process undertaken by a stakeholder working group, any linking of program theory to social science theory, and a formal consideration of any pending decisions and judgments informed by a review of program documents, a review of social science literature, and the program theory articulation process, as well as personal and professional experience and knowledge. The influences on evaluator theory are represented to the left in Figure 3.3 with the text detailing the influences on evaluator theory development.

Program documentation may include a written description of the program, mission statements, job descriptions or manuals detailing program operation procedures (Rossi, Freeman, and Lipsey, 1999). There may be documents relevant to legislative history to consider. In addition, any documents pertinent to program funding (e.g. grant applications, financial reports) may provide important insights into program processes. If any program theory has been articulated previously, even in an incomplete form, it should be uncovered and examined in a review of program documents.

The evaluator's hypotheses, in turn, may, at some level, influence the program theory explicated by a stakeholder working group, and the link of program theory to social science theory, and will certainly influence the evaluator's

thinking regarding any pending decisions and judgments to be made and the information needs to be addressed in the evaluation process. The Stakeholder Program Theory explication process normally would be facilitated by the evaluator and in return influence the evaluator's view of the program.

A review of pertinent social science literature is an important source of information that may help to support assumptions of the program theory and possibly raise questions regarding others. There are levels of theory that may be borrowed from the literature review. At the lower level there are theories developed and tested by social researchers in particular or well defined circumstances. Moving up the hierarchy we find social science theories particular to human behaviour and less influenced by circumstances (e.g. Ajzen and Fishbein's (1980) theory of reasoned action). Social science theory is particularly useful in the definition of program causative theory (Chen, 2004). For this reason, the social science literature is introduced to the stakeholder work group for consideration in their task of program theory articulation.

The selection of reviewed social science literature is influenced by what is unearthed in any searches undertaken, literature referenced in any program documents reviewed, relevant stakeholder expressions, and the evaluator's education, background and experience. The social science theory taken for the theory-based evaluation, unless one has been previously specified in the development of the program, typically is selected by the evaluator, in consultation with the stakeholders. In the event that a social science theory has been accessed for program development, the evaluator would normally encourage the stakeholder group to consider the theory with regard to appropriateness.

### Stakeholder Theory

Cummings (1997) considered program stakeholders to include the participants of the program, the staff of the program, those involved with similar programs,

policy-makers and the evaluators. There is generally a key group of stakeholders responsible for making decisions about a program's future. Additionally, there are a group of stakeholders whose lives are influenced by the program and any evaluation of that program also has the potential to influence their lives and therefore this group of stakeholders have a vested interest in the evaluation. These groups are all potential users of the evaluation information. As Cummings (1997) points out, all of the groups mentioned may be represented in the stakeholder group for a particular program evaluation, yet it is not possible to clearly determine the stakeholder group composition prior to beginning an evaluation study. Key stakeholders are identified primarily through a review of program documents (e.g. meeting notes, committee member lists) and in interviews with key program representatives.

In the same time frame that the evaluator begins undertaking a review of relevant program documentation, and a review of the social science literature, stakeholder groups knowledgeable about the program are identified and requested to participate in a working group for the purpose of theory development. The desired outcome is to represent their views of the program rationale, including the mechanisms through which they believe the program objectives can be attained. Essentially, it is a parallel view to that represented in the evaluator's theory.

In establishing a stakeholder working group for program theory articulation, a good representation of stakeholders with insight into program processes, activities and intended outcomes is crucial. Those responsible for the design and development of the program are important, as are those responsible for program implementation. Consideration of those responsible for the use of the evaluation information, and the purpose of the evaluation may help to guide the identification of key stakeholders.

*Program Theory Articulation.* The task of the stakeholder working group is to articulate the program theory from their perspective (Figure 3.3 represents the stakeholder program theory articulation process). In short, informed by relevant social science literature and relevant program documentation, the

evaluator facilitates the working group by means of brainstorming sessions to obtain the group's agreement on the underlying program theory including program activities, the program outcomes (short-term, intermediate and long-term) and the processes mediating between program activities and between program activities and program outcomes.

Once a clear program theory has been articulated, a graphic representation of the theory is presented to the stakeholder work group for verification and feedback. Revisions to the theory are negotiated. Clearly, the evaluator's consideration of the logic of the program will be influenced by this stakeholder work group articulation process.

The linking of the program theory developed with the stakeholders to pertinent social science theory is an elective step in the conduct of a theory-based evaluation, determined, for example, by the purpose of the evaluation, the evaluator's experience and preference, and the stakeholder's perception of the merits of tying the program theory to social science theory.

Social science theory is often used to suggest and articulate the causative mechanisms linking program activities to program outcomes. However, the use of established social science theory to delineate the assumptions regarding the mediating processes underlying the program, is a step that might not be taken in the theory-based evaluation. The case may be that the stakeholders are not interested in the higher order theoretical link. The evaluation may only be intended for program specific development. In such a case the evaluator may choose to not access social science theory, or may be unable to convince stakeholders of the value of social science theory to the evaluation process. The undertaking of a theory-based evaluation on the strength of the program theory originating with the stakeholders of the program, is a common path. Certainly, social science theory can be important for increasing the generalisability of the evaluation findings and enabling the determination of long-term outcomes (Chen, 1990). The use of social science theory to clarify and elucidate the underlying mediating mechanisms of the program process is of particular value. However, the reality is that a sound theory-based

evaluation may be, and often is, undertaken without accessing the social science theory.

In some cases experimental and conceptual work, undertaken in similar circumstances and context, identified in a literature review, is used to substantiate program theory. Evaluators may choose to use context relevant conceptual and experimental social science literature as opposed to grander, well-established social science theory to support and clarify the theory of the program to be evaluated.

### **Determine Needed Information**

Contact with stakeholders and a review of program documentation inform an understanding of the purpose of the evaluation, and the determination of decisions and judgments to be made regarding the program and possibly beyond the program, e.g. in relation to broader policy. The reason for undertaking an evaluation provides considerable insight into the needed information. This may be evident in the contracting of the evaluation. Through a review of relevant program documentation, conversation with stakeholders, and in the process of program theory explication undertaken with the stakeholder working group, the evaluator may gain insight into the purpose of the evaluation and any information the stakeholders have a particular interest in. A consideration of what information is most useful to which stakeholders will have already begun with the identification of the stakeholder groups.

The careful determination of information needs is relevant to the usefulness of the evaluation, whether a theory-based approach is adopted or not. However, the explicated program theory serves to highlight variables to be targeted in the evaluation design and serves as a guide to the information needed to assess program implementation, impact and the mediating variables. Furthermore, contextual and background variables may have a significant moderating effect on intended program processes and outcomes. A consideration of these variables in program design may provide useful insights into the influences



causing differences between the program as implemented and the program as intended.

#### OBTAINING, PROVIDING, UTILISING

The remaining three evaluation stages are undertaken in a manner similar to any non-theory-based evaluation. A representation of the theory-based evaluation stages is presented in Figure 3.3. The difference is that the theory of the program, carefully explicated in the delineating stage of the evaluation, guides the obtaining process, frames the providing process and influences the utilising process.

The information requirements which the evaluation must address are identified. The information needs are prioritised with reference to the purpose of the evaluation and the audience roles and requirements. In the development of measures to assess program activities, outcomes and mediating processes, a consideration of what is necessary, what is sufficient and what is optimal to produce the intended outcomes is essential to the generation of a basis for judging the intensity or strength of the program implemented (Chen, 1990a). Consultation with the stakeholder working group in the development of measurement specification is necessary. The consideration and measurement of inputs critical to program achievement of intended goals is crucial to the assessment of program implementation with regards to intended program activities. Critical inputs include the resources necessary for intended program implementation, and the mode of program delivery specifying intended program procedures, delineated in terms of duration, intensity and frequency (Sidani and Sechrest, 1999).

#### SUMMARY

The evaluation process model developed in this study is an initial step toward the establishment of non-theory-based and theory-based evaluation practice guidelines for the proposed study. The ability to determine the influence of the use of theory in an evaluation on the use of its information must be preceded

by a sound consideration of what theory use in an evaluation is, and how that takes place. The Theory-Based Evaluation Process Model (Figure 3.3) has been developed to detail this.

The assumption is that the use of the program theory to elucidate and to guide the collection, analysis, interpretation and documentation of the data increases the usefulness of the evaluation information (Chen and Rossi, 1981). Comparing the program as intended with the program as implemented, enables the consideration of discrepancies and facilitates consideration of strategies to adjust the program, if necessary. Information regarding the assumed processes mediating program activities and program outcomes, enables the consideration of the actualisation of these assumptions in program implementation. In essence, the insights gained from theory-based evaluation information regarding the program evaluated, should enhance the use of the information, as compared to a non-theory-based evaluation approach, and in particular black-box type of evaluation, both conceptually and instrumentally.

In the following section, a model developed with a view to understanding the variables with influence on the development of a generic program's theory, is elaborated in five stages. Versions of the model are presented with each stage. The final two versions of the program theory development model are concerned with the impact of the evaluation on the program's theory.

#### DESCRIPTIVE MODEL OF PROGRAM THEORY DEVELOPMENT

This section focuses on the development of a model representing a proposed temporal sequence of program theory development. The purpose is to further understand the processes and factors which influence the elaboration of the program theory. The influence of the layers of theory impacting on the program is considered, including relevant prescriptive theories, causative theory, implementation theory, and contextual considerations. The variables included are the identified need instigating the program, goals of the program, general and local specifications for the program, program reality and the impact of an evaluation on the program. The impact of the program theory on

the effect of the evaluation's information is then discussed. Each variable of the model is considered below, accompanied by a version of the model focusing on the pertinent variable.

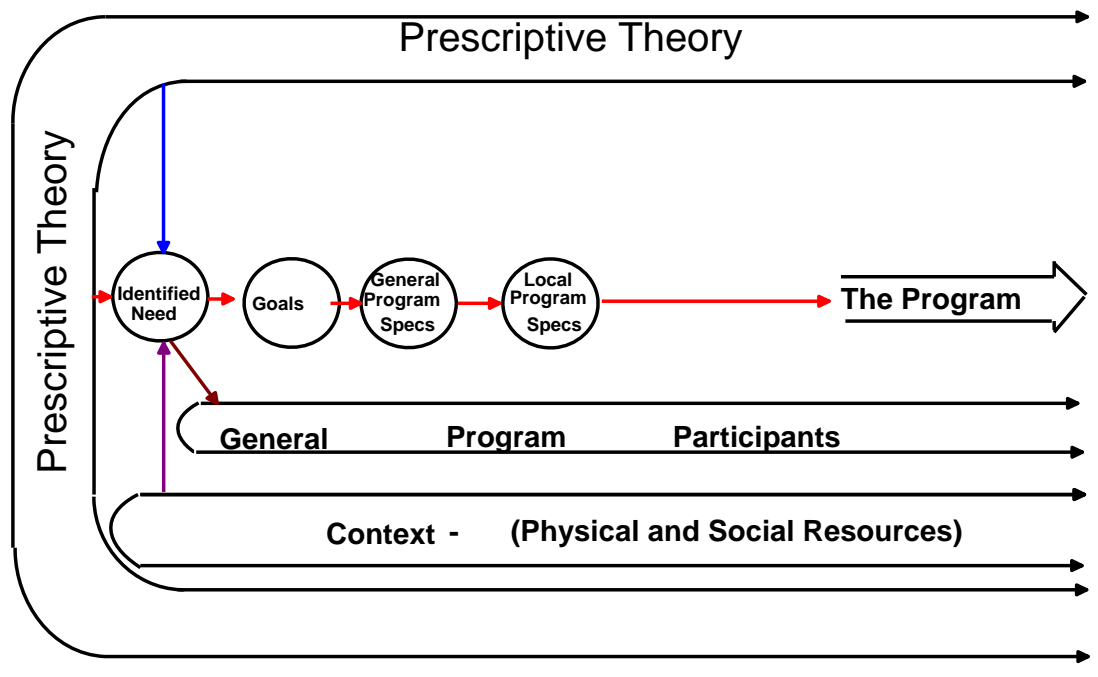
### **IDENTIFIED NEED**

Reflecting upon the temporal sequence of program theory development, it is apparent that program conceptualisation begins with the identification of a need (Figure 3.4). The identification of such a need, however, is based upon the prescriptive theories of the relevant society. In other words, their theories of the way things should be, are based primarily upon tacit knowledge embracing values, morals, beliefs and experience. Verbal and symbolic communications transmit these values and experiences and provide countenance for the establishment of societal norms; the often implicit rules and regulations we refer to in the course of our daily lives. Identification of a social problem rests upon the relevant prescriptive theories of society of the time, guided by the context of the need identification setting, including physical and social resources. Contextual factors that might influence the perception of a problem as a need, might include a change of governing body and pertinent policies, increased or reduced budgets, or a national disaster.

A concept central to a form of social intervention theory as considered by Lipsey (1997) is the 'criterion-threshold'. This threshold represents a point at which some condition is identified as a problem. It is the values and beliefs, the prescriptive theories of society, guided by the current contextual conditions surrounding a problematic condition that determine when a condition has passed the threshold from acceptable to non-acceptable, and thereby becomes a problem significant enough for society to direct resources towards its amelioration.

The identification of the need to be alleviated determines further components central to program development. The first is a conceptualisation of the general program participant population. The group affected by the need determined to

be a problem represents the general population from which the program participants will come.

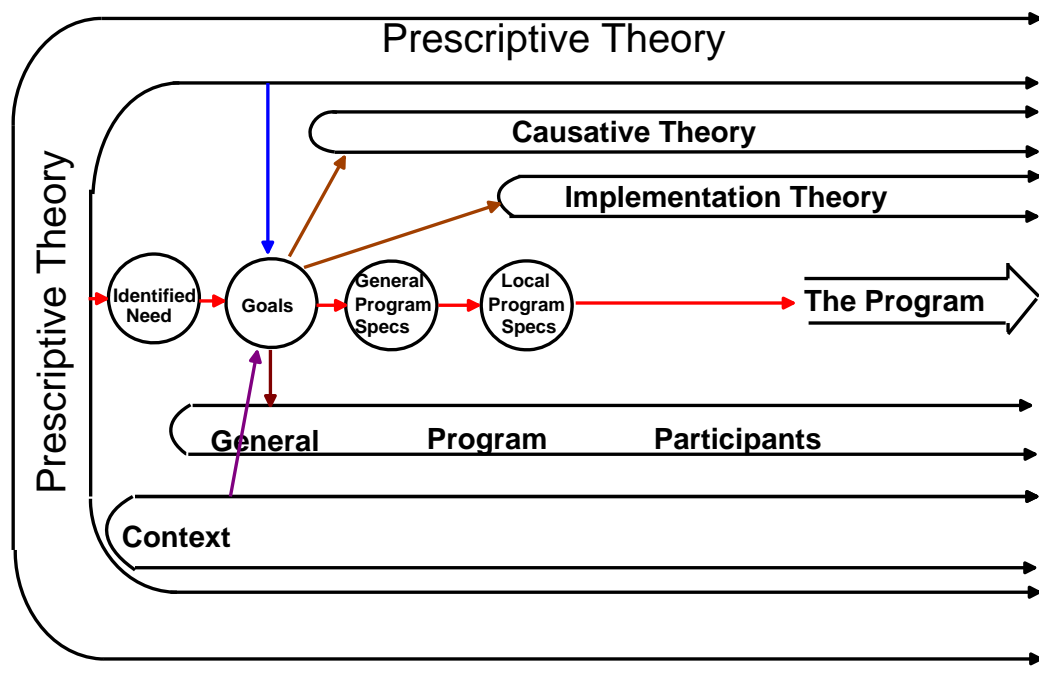


**Figure 3.4: Theory Development for a Program: Identified Need, General Program Participants**

## GOALS

Program goals (Figure 3.5) are the second consideration following from the perception of a need to be alleviated. The determination of the goals to be addressed by a program initiated to meet an identified need, is a process resting upon the prescriptive theories of key stakeholders including, among others, policy-makers at the national, state, and local levels.

Ideally, policy-makers and program developers might consider the prescriptive theories of the community, special interest groups, and possibly the population targeted by the program under development. Furthermore, the consideration of program goals further delineates the general program participant population, and is an initial step in the development of a program's implementation theory, or the theory of program operation linking the program processes together.



**Figure 3.5:** Theory Development for a Program: Program Goals, General Program Participants, Causative Theory, and Implementation Theory

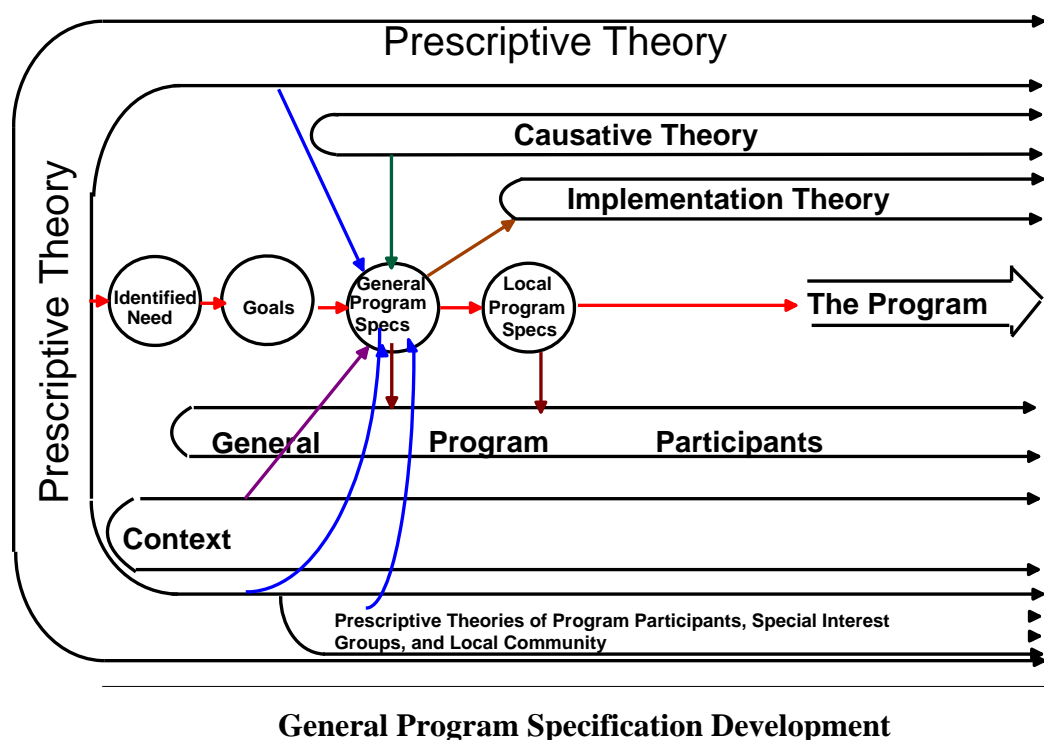
Once the goals have been determined, a strategy for achieving them must be generated. With the intended goals in mind, the prescriptive theories of key stakeholders grounded in the context of the situation, the physical and social environment, including the availability of resources, inform the selection of a satisfactory causative theory. Stakeholders' prescriptive theories (for example a high regard for the psycho-social model over the medical model for treating drug dependency) directs their choice of a strategy for achieving intended goals, yet the availability of resources to support the strategy determines the viability of the causative theory. A lack of available resources may negate the adoption of a particular causative theory and necessitate the selection of another reasonable, and valued, avenue for achieving the desired goals. Thus, the selected causative theory contributes to the development of the program's implementation theory.

## GENERAL PROGRAM SPECIFICATIONS

The general program specifications tend to emerge in the process of considering possible causative theories (Figure 3.6). In the course of weighing possible causative theories, policy-makers may implicitly and explicitly consider prescriptive theories; their own and those of the community, special interest groups, program practitioners and possible program participants.

Additionally, the physical and social environment (resources) or context within which the program is developed, informs the selection of a program's causative theory and the general program specifications. Certainly, the scale of the program, the size of the program participation population, and the level of training of program practitioners employed to deliver the program, are all a function of the resources made available to the program. Again, general program specifications contribute to the development of the program's implementation theory.

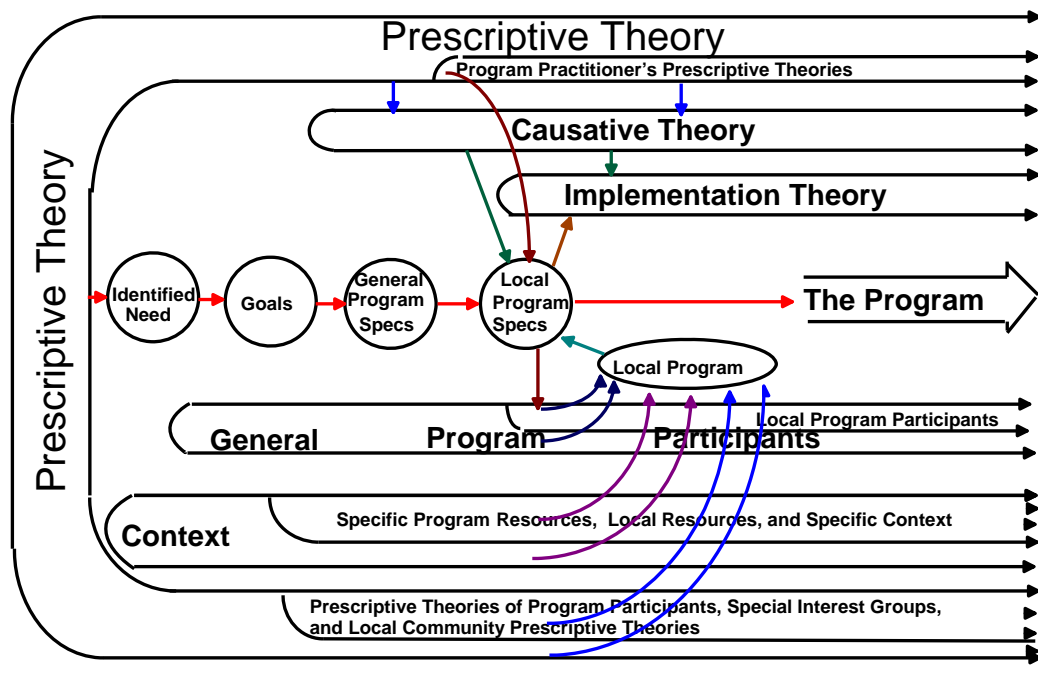
**Figure 3.6: Theory Development for a Program:**



## **LOCAL PROGRAM SPECIFICATIONS**

In consideration of a program's implementation, local program specifications are determined through an interaction of factors (Figure 3.7). All of the factors influencing the program's development to this point continue to be represented, however, they are considered in terms of their components, or local program considerations. Implementation of the general program specifications, guided by contextual circumstances of the program including the local and specific resource availability, the specific contextual conditions into which the program is introduced, and prescriptive theories of those with an interest in the program, all contribute to determine the local program specifications, and, in turn, the local program participants. The determination of the local program participant group influences the program's implementation theory.

Physical and social resources include the resources specific to the program, the local resources and context specific resources. Resources specific to the program include resources earmarked for the program, such as funding made available to the program for implementation. Local resources are those available in the local community which might be recruited for program implementation. For example, the availability of a community hall as a venue for the program to take place might be a useful resource available in the local community. The specific context resources into which a program is introduced include physical and social elements endemic to the local program specification. For example, a program aimed at increasing the school attendance rates of high school students from families of low socio-economic status would be implemented in a school with a high number of students from disadvantaged families, rather than in a school serving a cluster of suburbs with families of middle to high socio-economic status.



**Figure 3.7: Theory Development for a Program: Local Program Specification Development**

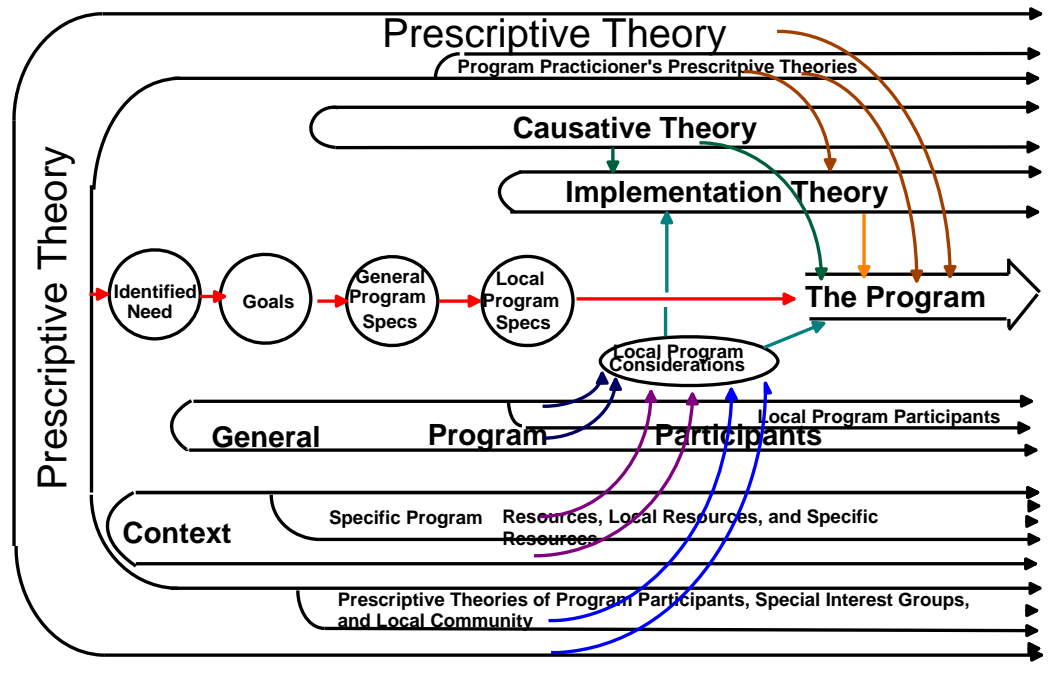
The prescriptive theories of six groups of stakeholders influence the local program specifications; policy-makers, program developers, program practitioners, program participants, special interest groups, and the local community. The prescriptive theories of policy-makers and program developers, reflecting the norms, values and beliefs of society in general, continue to have an influence, yet they are subordinate to the prescriptive theories of the program practitioners; the people responsible for the implementation of the program. There is an expectation that the prescriptive theories of program practitioners will, to a great extent, align with those of relevant policy-makers and program developers due to the selection process leading to their employment. However, they also bring their individual beliefs, values, experiences and training to the program which is implicitly reflected in their practices. Their interpretation of the program's theory is guided by their individualized prescriptive theories, which come into play when the program is implemented, further guiding the delineation of the local program specifications.



There are further ways in which the prescriptive theories of program participants, special interest groups and the local community shape local program specifications. For example, if a program does not take into account the prescriptive theories of the local participants, e.g. in relation to cultural values, there is a likelihood of alienating the group the program is endeavouring to change. Additionally, special interest groups, for example the parents of children with a physical or mental disability, or the carers of those with a debilitating illness, are also influential and will have opinions about the way things should and should not be done under a program which targets a group they have a special interest in, or experience of. Finally, the prescriptive theories of the local community in which a program is introduced influence local program specifications. For example, a local community may be resistant to the placement of a halfway house for the purpose of reintroducing juvenile criminal offenders into society, in close proximity to a primary school. In such cases, the likelihood of that program being a success is diminished. In fact, it may actually be counter-productive.

### **PROGRAM REALITY**

The application of local program specifications, again negotiated through local program considerations of the prescriptive theories of the relevant society (policy-makers and program developers), program practitioners, program participants, special interest groups, and the local community, the general, local and specific resources available to the program, and the general and specific context into which a program is introduced, all significantly influence the program's implementation theory. It is only upon implementation that all the factors intermingle and the program reality is brought about. Figure 3.8 represents a composite graphical representation of project theory development, tracing the influences from the identification of a need to be alleviated to the operationalisation of a program's reality.



**Figure 3.8: Theory Development for a Program: The Program**

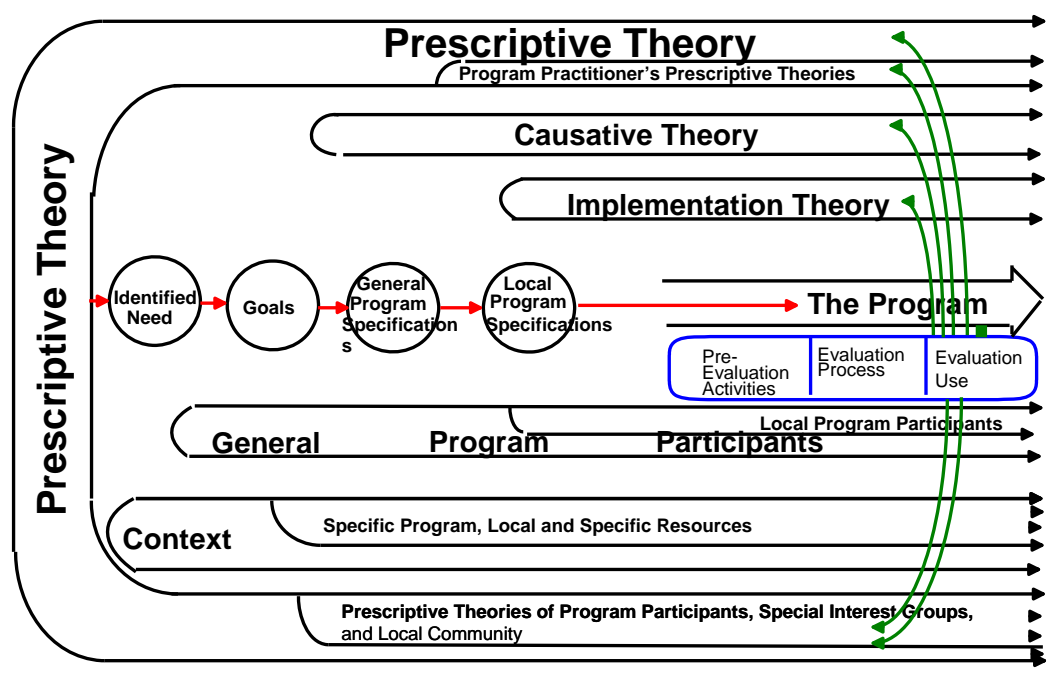
## THE EVALUATION INFORMATION USE ISSUE

Evaluation information utilization has been an area of interest in program evaluation since its inception as we know it today (Johnson, 1998) and represents a significant factor in the model delineating the theory development of a program detailed here. In 1966 Weiss presented a paper to the American Sociological Association in which she stated that the most common complaint of evaluators was that the findings of the program evaluations they undertook were not utilised (Weiss, 1972). Over the past three decades many authors have contributed to the wealth of program evaluation utilisation literature (e.g. Alkin et al., 1979; Guba, 1969; Cousins and Earl, 1992; Cousins and Leithwood, 1986, 1993; Greene, 1988; Leviton and Hughes, 1981; Johnson, 1998; Shadish, Cook and Leviton, 1991; Shea et al., 1995; Shulha and ousins, 1997) and to the slow-paced, yet balanced, development of understanding of the use of program evaluation information (Johnson, 1998). The literature supporting the concern regarding the evaluation information use issue has been presented in Chapter 2 of this dissertation.

## THEORY DEVELOPMENT FOR A PROJECT: EVALUATION USE

Weiss's (1972) declaration that the fundamental rationale of program evaluation was the provision of information for program decisions was a stand many authors of the time agreed with (e.g. Caro, 1971; Scriven, 1967). Literature detailing instrumental and conceptual use (e.g. Rich, 1977; Weiss, 1972), process use (e.g. Greene 1988; Patton, 1997), symbolic use (Shula and Cousins, 1997), persuasive use (Leviton and Hughes, 1981) and strategic use (Cummings, 1997) presented in Chapter 2 of this dissertation.

In the model of theory development for a project, the components in the model represent areas for evaluation utilisation. Conceptual use of evaluation information (Figure 3.9), including strategic use and process use, will be initially reflected in alterations in the abstract conceptual sub-theory components of the model. These include the prescriptive sub-theories of key program stakeholders such as program practitioners, the local community, and special interest groups, as well as local and general project participants, and the way key stakeholders think about the project's causative sub-theory and implementation sub-theory.



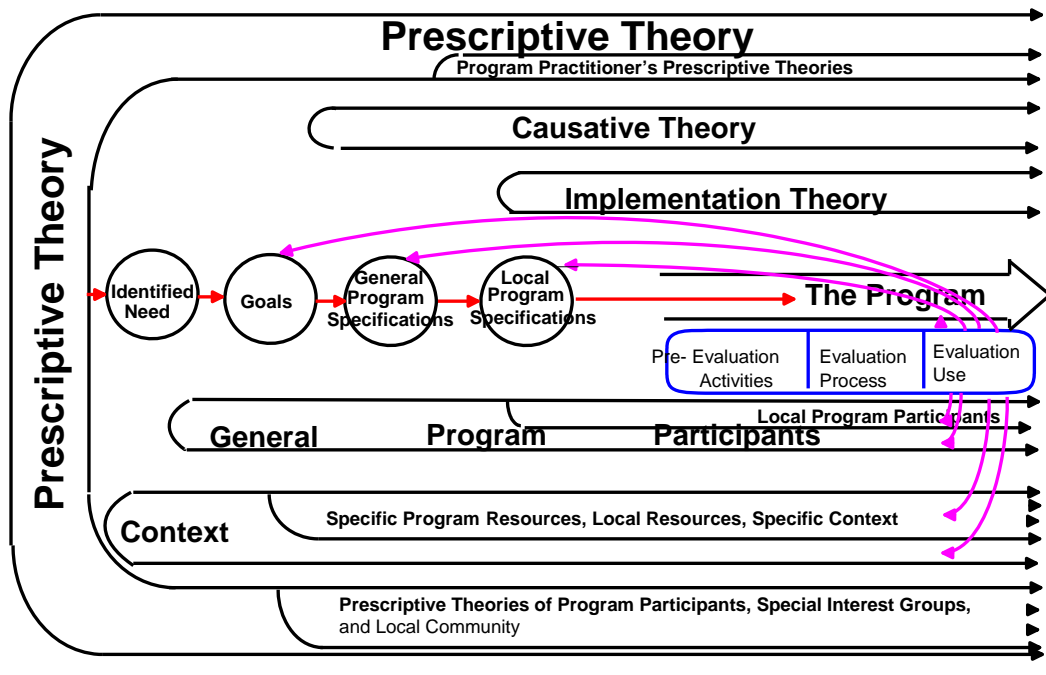
**Figure 3.9: Theory Development for a Program: Project Evaluation Conceptual Utilisation**

Following on from conceptual use, or the reconsideration of prescriptive theories relevant to the project, the project's implementation theory or causative theory recognised as merited by stakeholders in a decision-making capacity, are likely to influence decisions made regarding the local program specifications, general program specifications, general or local program participants, and even the goals of the project, if the evaluation finds the project's intended goals too difficult to achieve or inappropriate for the project actualised (Figure 3.10). Furthermore, instrumental use may occur in a redistribution of resources. If the program is successful in achieving its intended *goals*: if it proves itself, then the specific, local and general resources appropriated to the program may be increased. Alternatively, if the program is not successful, or not as successful as intended, resources initially made available for its implementation may be diverted to proposed programs, or to programs already operating successfully.

The influence of any use of the evaluation, conceptual or instrumental, results in an altered program reality. The effect of these changes is weighed by stakeholders in an ongoing informal evaluation process which the evaluation has become a part of, and thereby the theory development cycle continues.

#### **THEORY-BASED EVALUATION USE AND THE PROJECT THEORY DEVELOPMENT MODEL**

As outlined earlier in this study, the merits of undertaking a theory-based program evaluation rest upon the increased usefulness of the evaluation's information over a non-theory-based program evaluation. It is expected that a theory-based evaluation addressing the mechanisms by which the project's outcomes occur, in-other-words the program's causative theory, will have more impact on thinking regarding the causative theoretical components of the model than a non-theory-based evaluation in which measures of the program's causative mechanisms are absent. Logically, increased conceptual use results in an increased capacity for instrumental use to occur.



**Figure 3.10: Theory Development for a Program:  
Project Evaluation Instrumental Utilisation**

A second considered merit of undertaking a theory-based evaluation is the increased generalisability of the information, dependent on the manner in which the causative theory is elucidated. If this is the case, the use of theory-based evaluation information should be reflected in changes in prescriptive theories held by stakeholders regarding the way events should occur. For example, if a theory-based evaluation shows a particular causative theory to be ineffective with a particular group, the relevant stakeholders can absorb this information into the prescriptive theories they hold regarding the processes of such programs and possibly apply it to later programs and other relevant experiences.

With regards to the implementation sub-theory component, it is expected that increased causative theory information will result in more carefully conceptually delineated implementation theory with an awareness of program mediators and moderators. Non-theory-based evaluation commonly presents information regarding the program's implementation theory from an external point of view in a description of how the program components are linked together, in other words, how the program works. A theory-based evaluation

incorporating measures of the mechanisms by which the program's intended outcomes are brought about, is expected to provide more powerful information relevant to a program's implementation theory due to the increased information regarding the layers of the program's operation. Therefore, the theory-based evaluation information should result in an increased awareness (conceptual use) of program operation over that occurring as a result of information from a non-theory-based evaluation. The increased conceptual use of theory-based evaluation over non-theory-based evaluation would be expected to offer a greater likelihood of instrumental use of theory-based evaluation information.

## **SUMMARY**

This section has focused on the variables upon which a theory of a program are formed, and the effect of an evaluation of the program on the theory. A model representing the temporal sequence of program theory development has been elaborated. The purpose of this model is to further understanding of the processes and influences important to the elaboration of the program theory and the effect of the evaluation's information. The variables considered have included the identified need instigating program development, the goals of the program, the general and local specifications for the program, the program as realized in practice and the impact of an evaluation on the program. The influence of the layers of theory with an influence on the program has been considered, including relevant prescriptive theories, causative theory, implementation theory, and contextual considerations.

The following section focuses on the theory-based data collection matrix developed for this study. The matrix further develops concepts presented in Stake's 'Descriptive Matrix' (Stake, 1967, p. 529), reproduced here as Figure 3.11. The purpose of the matrix is to guide the collection of data relevant to a theory-based evaluation in order to judge the extent of theory use in an evaluation study.

## THEORY-BASED DATA COLLECTION MATRIX

In Stake's 1967 paper 'The Countenance of Educational Evaluation' he advocated full description of the program evaluated. In light of the strong experimental persuasion present in program evaluation practice at that time, the qualitative approach in evaluation was a controversial and little explored path. Stake proposed a matrix for describing a program in terms of intents and observations in three areas: antecedents, transactions and outcomes (see Figure 3.11). The concept and form of the matrix presented here, to be considered as a guide to data collection in the conduct of theory-based evaluation, was initially guided conceptually by the 'Descriptive Matrix' (Stake, 1967; p. 529) portion of Stake's data collection matrix. The present matrix, however, is fundamentally different from that proposed by Stake (1967) in that it incorporates description of program theory, where in Stake's matrix he considers a statement of program rationale necessary for a consummate evaluation, but emphasises that it be considered separately from the data collection matrix.

Similar to Stake's Descriptive Matrix, the matrix developed for the present work (Table 3.1) considers program data in terms of the Intended Program, the Observed Program, and Evaluation Information Use. These program states are separated by the actualisation of Program Implementation. Each program state considers three elements of the program, the Program Outcomes, the Causative Theory, and the Implementation Theory. The Data Collection Matrix developed here includes the causative theory of the program, the Implementation Theory of the program and the Program Outcomes as important information collection focuses.

The Evaluation Implementation enables the consideration of Evaluation Information Use. With regard to each of the three program elements, the collection of information about the use of the evaluation information is a step beyond the Descriptive Matrix of Stake, yet pertinent to the basis of the investigation proposed here. The matrix components are detailed below, followed by a discussion of the logic of matrix information analysis.

**Table 3.1: Theory-Based Evaluation Data Collection Matrix**

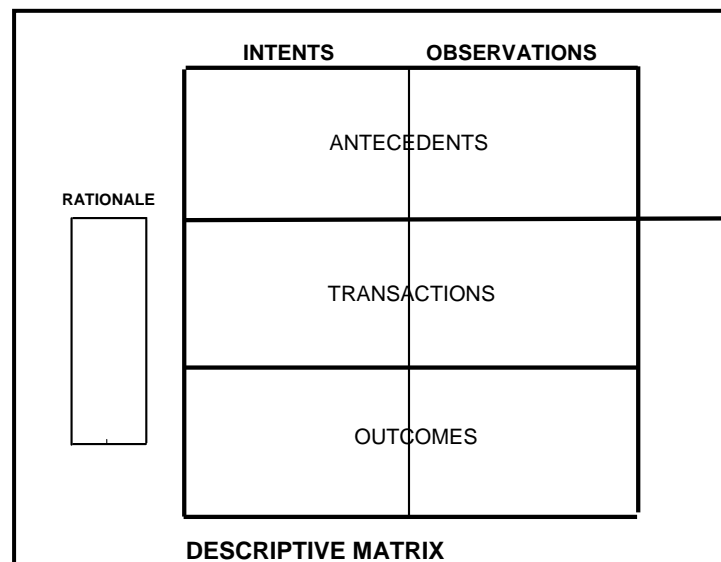
Identified Problem	Intended Program (Expected Context)		Observed Program (Observed Context)		Evaluation Information Use
Outcomes	<ul style="list-style-type: none"> <li>♦ What is the program expected to achieve? <ul style="list-style-type: none"> <li>▪ Immediate Goals</li> <li>▪ Intermediate Goals</li> <li>▪ Long-term Goals</li> </ul> </li> </ul>		<ul style="list-style-type: none"> <li>♦ What has the program achieved? <ul style="list-style-type: none"> <li>▪ Intended Outcomes</li> <li>▪ Unintended Outcomes</li> </ul> </li> </ul>		<ul style="list-style-type: none"> <li>♦ How was outcome information used conceptually and instrumentally?</li> <li>♦ What factors influenced the information use and in what ways?</li> <li>♦ How did the evaluation processes and practices influence the use of the evaluation and its information?</li> </ul>
Causative Theory	<ul style="list-style-type: none"> <li>♦ How will these outcomes be achieved?</li> <li>♦ What are the intended causal mechanisms that are expected to mediate the program processes and practices and the program outcome.</li> </ul>	Program Implementation	<ul style="list-style-type: none"> <li>♦ Were the links between the program processes and practices and the intended program outcomes actualised?</li> </ul>	Evaluation Implementation	<ul style="list-style-type: none"> <li>♦ How was program causative theory information use conceptually and instrumentally?</li> <li>♦ What factors influenced the information use and in what ways?</li> </ul>
Implementation Theory	<ul style="list-style-type: none"> <li>♦ What resources are necessary for intended program delivery?</li> <li>♦ What are the processes and practices by which these outcomes are expected to be achieved?</li> <li>♦ What variables are expected to affect these processes and practices?</li> <li>♦ What variables are expected to moderate the effect of these processes and practices?</li> </ul>		<ul style="list-style-type: none"> <li>♦ What were the program processes and practices by which the outcomes were achieved?</li> <li>♦ What variables affected these processes and practices?</li> <li>♦ What variables moderated in the effect of these processes and practices?</li> </ul>		<ul style="list-style-type: none"> <li>♦ How was program implementation theory information used, conceptually and instrumentally?</li> <li>♦ “What factors influenced this use and in what ways?”</li> </ul>



## INTENDED PROGRAM, OBSERVED PROGRAM AND EVALUATION INFORMATION USE

### INTENDED PROGRAM

The Intended Program column of the matrix is concerned with the collection of information regarding the program as it was originally intended by program developers and relevant stakeholders, including the inputs of the program. The Intended Program, a result of numerous weighted considerations and decisions regarding the program outcomes, causative theory and implementation theory, is an expression of the prescriptive theory of those responsible for program development.



**Figure 3.11: Descriptive Matrix presented by Stake (1967; p. 529) as a portion of his Data Collection Matrix.**

Although prescriptive theory is commonly assumed by program developers and relevant stakeholders, and usually not stated explicitly or examined, it determines the rationale and justification for the program structure and activities. It guides program planning, development and implementation (Chen, 1990b). When deciding the form and course of the program, in terms of its nature, duration and intensity, numerous value judgments are made based

on the beliefs, opinions and attitudes of those making the decisions and these are considered with regard for optimal program effectiveness in the achievement of program outcomes. Those responsible for developing a program consider the nature, form and duration of the practices they value in light of how effective these practices are in achieving program outcomes. In short, intended program data represents the prescriptive theory of program developers including a value of program effectiveness. The expected context of the program is yet another consideration with implications for program decisions.

Those responsible for the development of a program consider its implementation in a particular context which is determined predominantly by where and how the program occurs. The placement of the program in a particular context allows decision-makers to address situational variables that may influence the program's processes, practices and outcomes. The information represented in the Intended Program column of the matrix includes the context into which the program is expected to be introduced.

#### OBSERVED PROGRAM

The Observed Program is the intended program post-implementation. Were the inputs to the program as intended? Upon implementation of the Intended Program, situations and problems may arise that have not been envisaged by the program developers. Discrepancies between intended program context and observed program context may lead to changes to maximize the effectiveness of the program in the context within which the program is actualised. Additionally, although the prescriptive theory of program developers is carried through into the program as observed, with implementation comes the influence of the prescriptive theories of other program stakeholders, primarily program practitioners, but also the program participants, special interest groups and the local community. The result of the influence of these additional stakeholder prescriptive theories may be alterations in program practices to bring them in line with the prescriptive theories of the prevailing stakeholder

groups. Once again, decisions regarding program processes and practices are made with a view to program effectiveness. Program effectiveness is a responsibility handed down to program practitioners responsible for implementing a program.

#### EVALUATION INFORMATION USE, THEORY-BASED EVALUATION DESIGN: CONGRUENCE AND CONTINGENCY

The proposed matrix provides a focus for the consideration of the congruence between the Intended Program and the Observed Program. The congruence of the Intended Program Outcomes, Intended Causative Theory and Intended Implementation Theory with their observed counterparts does not test directly the program theory, but rather informs whether what was intended did occur or not, and possibly some insight as to why. A consideration of congruency between program activities and program outcomes provides information regarding the extent to which the program theory is practicable.

It is important to recognise that the development of the intended program outcomes, causative theory and implementation theory is a dynamic interactive process, difficult to capture diagrammatically, in which decisions are made and then renegotiated in light of further considerations. However, like Stake's (1967) Descriptive Matrix, the expected sequences tying the Intended Outcomes, Intended Causative Program Theory, and Intended Implementation Theory into a comprehensive program theory must be taken into account and assessed for soundness through logical reasoning. Alternatively, links between Observed Program Outcomes, Observed Causative Theory and Observed Implementation Theory may be assessed empirically, in addition to examination through reasoning based on past experience and accumulated knowledge.

### **Outcomes**

The outcomes of the program are the changes which occur as a result of program implementation. The focus of this section is on Intended Program Outcomes and Observed Program Outcomes. Intended Program Outcomes are considered in light of the identified need or problem to be alleviated by the program and are the goals the program is undertaken to achieve. They are important to stakeholders because they are the intended consequence of the program, immediate, intermediate and distal (Stake, 1967). Intended Outcomes guide stakeholders' program related activities, are a consideration in decisions made regarding resource allocation and are commonly used as a standard to assess program effectiveness (Chen, 1990b). At this stage of program development, some effort may also be made to ascertain unintended program outcomes, though such outcomes most probably will not be fully realized until post-implementation.

The discrepancy between Intended Program Outcomes and Observed Program Outcomes is often the basis upon which program success is measured. However, the influence of context and relevant prescriptive theories may moderate the assessment. Immediate, intermediate, and long-term Observed Program Outcomes may differ from intended outcomes due, for example, to program changes instigated by implementation of the program in a real context, as opposed to an abstract context, or the influence of regnant stakeholders' prescriptive theories.

### **Causative Theory**

In the matrix, the Causative Theory of the program, or the program rationale, is placed in the centre of the data collection process. Once the Intended Outcomes of the program have been determined, decisions must be made regarding the underlying mechanisms that mediate the program's Intended Implementation Theory and the Intended Outcomes. Possible causative

theories to be adopted to achieve the Intended Program Outcomes are identified. A process of weighing the choices in terms of the intended context, the prescriptive theories of program developers, and optimal program effectiveness results in the selection of the Intended Causative Theory to be implemented.

The Observed Causative Theory of the program, post-implementation, should ideally reflect the Intended Causative Theory. However, the practicalities of program implementation, the context within which the program is actualised, and the influence of the prescriptive theories of dominant stakeholders are examples of factors which may encourage adjustment or alteration of the causative theory. The influence of these elements may sway program practices to the point where the intended causative theory is defunct. The determination of a discrepancy between causative theory observed and causative theory intended may have important implications for Observed Program Outcomes.

### **Implementation Theory**

The Implementation Theory of the program is the logic including and linking the resources and activities of the program. Implementation Theory tests the assumption that if a program is put into action, as intended, the intended outcomes will be brought about. Simply, it specifies how the program is to be put into action. Decisions made by program developers and other relevant stakeholders regarding the Intended Implementation Theory implicitly, though often not explicitly, involve consideration of the relevant prescriptive theories, optimal program effectiveness, the intended context of the program, the intended program outcomes, and the chosen causative theory. Actual implementation of the program is the test of the intended implementation theory that results from the interchange of these factors.

What were the actual program processes and practices? The practicalities of conducting a program in a context that may or may not resemble the intended program context, and the influence of the prevailing stakeholders' prescriptive theories, may result in a discrepancy between the Intended Program

Implementation Theory and the Observed Implementation Theory. The impact of these changes on the Observed Causative Theory and Observed Outcomes of the program may be of critical importance to program understanding.

#### AN EXAMPLE

An example of the use of the Data Collection Matrix using the theory-based evaluation of the Garbage Reduction Program in Taiwan, undertaken by Chen, Wang, and Lin (1997), is presented below in Table 3.2. A more detailed explanation of the classification of the program theory use of this particular evaluation is considered later in this chapter.

The Data Collection Matrix is a tool used to ensure that information pertinent to each of the elements of a theory-based evaluation considered critical in the study proposed in this dissertation is collected in a consistent manner. In sound theory-based evaluation, clear statements regarding the intended program outcomes, intended causative theory and intended implementation theory should be present based on a cogent investigative process undertaken to articulate the program's theory. This process will most likely have included a review of relevant program documentation, a review of pertinent social science literature and a program theory articulation exercise undertaken with program stakeholders.

Based on the intended program theory articulated, strategies and measures are developed to assess the observed program and determine the extent to which it is congruent with the intended program in terms of outcomes, causative theory and implementation theory. In addition, the contingency of the program processes and practices is assessed. In the example in Table 3.2, the outcomes of the program were not contingent on the implementation of the program because the causative theory upon which the program is based is inaccurate. As indicated in Table 3.2, the underlying causative theory of the program relied on the residents being inconvenienced by the unpleasant odour of having to store garbage in their home to motivate them to reduce garbage

output. In fact, the odour of garbage stored in their homes; the mechanism, did not move them to act. The residents simply endured the odour.

**Table 3.2: A Use Example of the Data Collection Matrix for the Evaluation of the Garbage Reduction Program in Taiwan, (Chen, Wang, and Lin (1997))**

<b>Problem-</b> Rapid Accumulation of Garbage	<b>Intended Program</b>  <b>Context- High (Taiwan Island)</b>		<b>Observed Program</b> <b>Context- High (Taiwan Island)</b>		<b>Evaluation Information Use</b>
<b>Outcomes</b>	<ul style="list-style-type: none"> <li>◆ Residents reduce personal garbage output on Tuesdays.</li> <li>◆ Residents reduce output on other days.</li> </ul>	<b>Program Implementation</b>	<ul style="list-style-type: none"> <li>◆ No impact on reducing garbage volume.</li> </ul>	<b>Evaluation Implementation</b>	<ul style="list-style-type: none"> <li>◆ How was the evaluation information used conceptually and instrumentally?</li> <li>◆ What factors influenced the use of information and in what ways?</li> </ul>
<b>Causative Theory</b>	<ul style="list-style-type: none"> <li>◆ Residents experience inconvenience of storing garbage.</li> <li>◆ Residents experience odour of storing garbage.</li> </ul>		<ul style="list-style-type: none"> <li>◆ Residents neither felt inconvenience in storing at home on Tuesdays or unpleasant odour.</li> </ul>		<ul style="list-style-type: none"> <li>◆ How was the evaluation causative theory information used, conceptually and instrumentally?</li> <li>◆ What factors influenced that use and in what ways?</li> </ul>
<b>Implementation Theory</b>	<ul style="list-style-type: none"> <li>◆ Inform residents of media campaign, mail, banners.</li> <li>◆ Implement program- No garbage collection on Tuesdays.</li> <li>◆ Monitor curb side dumping of garbage on Tuesdays.</li> </ul>		<ul style="list-style-type: none"> <li>◆ Site visits by evaluators to monitor program.</li> <li>◆ Program was determined to have been implemented as intended.</li> </ul>		<ul style="list-style-type: none"> <li>◆ How was the implementation information used, conceptually and instrumentally?</li> <li>◆ What factors influenced that use and in what ways?</li> </ul>

Unfortunately, this particular example provides little information regarding the use of the evaluation information beyond the expectation that it would be used at some stage to inform the development of future garbage reduction programs. Evaluation studies do not typically extend to the investigation of the use of their information. However, it is a step that will be taken in the investigation proposed here. Assessment of the use of the evaluation information will address the use of outcome information, causative theory information and implementation theory information. In addition, how the information was used, either conceptually or instrumentally, and the factors that influenced that use, including evaluation characteristics, evaluator characteristics, and stakeholder characteristics will be considered. The Data Collection Matrix encourages the collection of consistent information across evaluations, with a view to comparing the use of the evaluation information and the influences on that use.

## **SUMMARY**

The data collection matrices (Table 3.1) has been developed to highlight the information to be considered in the investigation of a theory-based evaluation and the use of its information, and in the investigation of a non-theory-based evaluation and the use of its information. The following section turns to the consideration of the model to be developed and tested in this study. The model which has been developed is based on the Theory-Based Evaluation Process Stages Model (Figure 3.3).

## **THE FORM OF THE MODEL TO BE TESTED**

The Core Model to be tested for this study is the focus of this section. The Core Model includes the primary factors found to contribute to the utilisation of evaluation information, identified in relevant empirical literature (Alkin, 1985; Cousins and Leithwood, 1986, 1993; Cummings, 1997; Leviton and Hughes, 1981; Hudson-Mabbs, 1993; Pawson and Tilley, 1997; Preskill and Torres, 1997). The program theory factor dealing with the influence of the use



of program theory in the evaluation (causative program theory and implementation program theory) on the use of the evaluation information, is a key predictor variable that is the focus of this study and is hence included in the model. Four additional latent predictor variables are included in the model. In addition, four separate outcome variables have been included, therefore four separate versions of the Core Model, one for each outcome variable, has been included.

The section below focuses on the Core Model, and details each observed variable and the latent variable which they indicate. Furthermore, the four latent outcome variables are detailed. A generic template of the Core Model is presented. The second half of the section focuses on a further elaborated version of the Core Model, including a further four latent predictor variables. Although previous research reported in the evaluation literature support the model hypothesised for this dissertation, this investigation is an initial exploration into a number of different areas which have not previously been investigated.

#### **LATENT VARIABLES OF THE CORE MODEL TO BE TESTED**

The Core Model of the study includes a common set of five latent variables, predicting in turn each of the four outcome variables. The five latent predictor variables are *Program Theory*, *Evaluation Study Characteristics*, *Process Use*, *Commitment to Study (Pre)*, and *Commitment to Study (Post)*. The four outcome variables are *Influence of Study Findings*, *Influence of Involvement in Elaboration of Program Theory*, *Use of Program Theory in the Final Report*, and *Importance/Likelihood of Information Use*. Each of the variables is summarised below. Diagrams representing versions of the Core Model (Figure 3.12) to be tested have also been included.

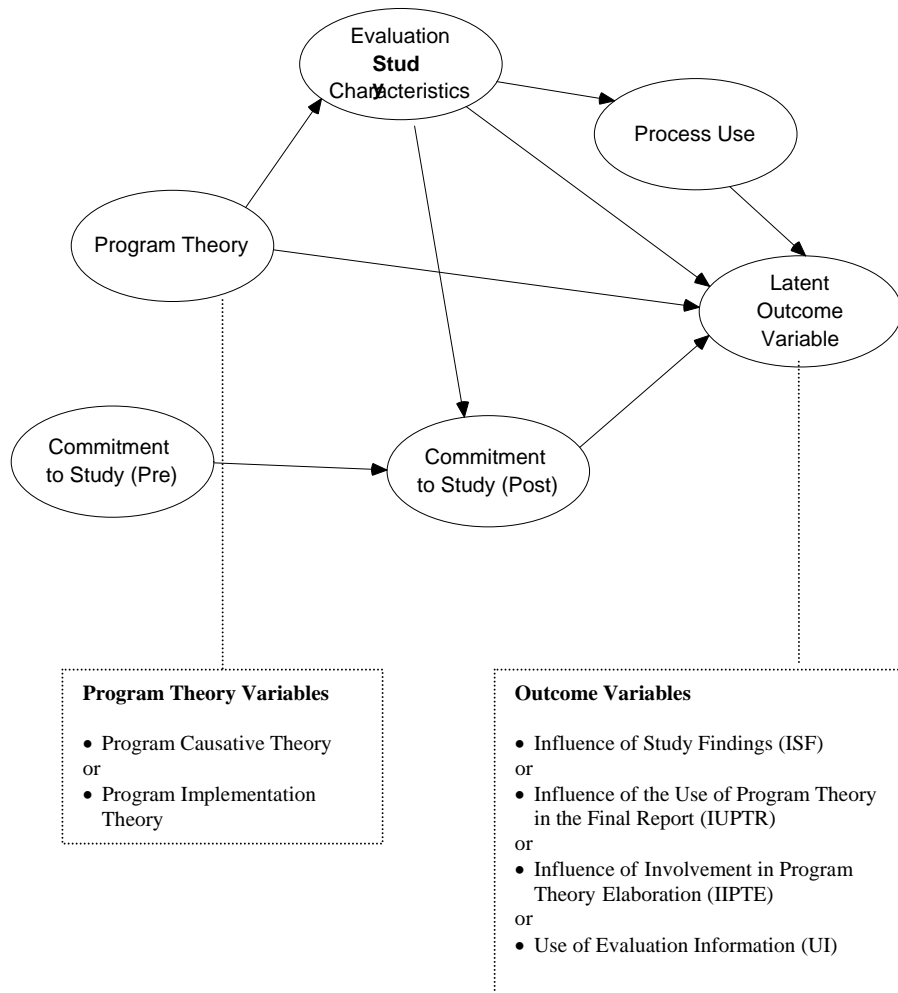
## PROGRAM THEORY

According to a number of authors, the use of a program's theory in an evaluation undertaking should result in an increased usefulness of the evaluation information. Implementation theory and causative theory, the two types of program theory discussed previously in this chapter, are the focus of the program theory Classification Matrix developed in the following section (Bickman, 1987; Chen, 1990b; Chen, 2003; Pawson and Tilley, 1997; Petrosino, 2000; Rogers et al., 2000; Scheirer, 1987; Suchman, 1967). In the testing of the Core Model these two types of program theory are considered in turn. The two theory types are components of the generalised *Program Theory* latent variable included in the Core Model. Thus, there are two versions of the Core Model, one incorporating *Program Implementation Theory* and the other *Program Causative Theory*. A correlation of the two types indicated they were not correlated (*Pearson Correlation* = -0.15 (*p-value. (2-tailed)*) = 0.485). The data used for the calculation was taken from the 21 studies included in the Program Theory Classification Matrix developed for this study. The Program Theory Classification Matrix classifies the type and level of program theory use in each evaluation study. Details regarding the process of program theory use classification for each program are presented later in this chapter. Based on the results of the correlation, a decision was made to develop two versions of the reduced model, one incorporating *Program Implementation Theory* and the other *Program Causative Theory*. A reduced structural model version, including the two program theory options, is represented in Figure 3.12.

## EVALUATION STUDY CHARACTERISTICS

The perceptions of stakeholders regarding certain characteristics of the evaluation study have been found to have an influence on the use of the evaluation information. The characteristics found to have influence include evaluation quality, relevance and timeliness of the findings, credibility, communication quality, approach of the evaluation team practice and environment of the evaluation (Alkin, 1985; Cousins and Leithwood, 1986, 1993; Cummings, 1997; Leviton and Hughes, 1981; Hudson-Mabbs, 1993;

Pawson and Tilley, 1997; Preskill and Torres, 1997). A more thorough review of the literature relevant to each has been undertaken in Chapter II. The *Evaluation Study Characteristics* latent variable is an important variable in the Core Model of this study. It is indicated by five observed variables, each measured by a scale. The observed variables are Study Characteristics 1 (SC1), Study Characteristics 2 (SC2), Evaluation Study Environment (ESE), and Evaluation Team Characteristics (ETC). Each observed variable is measured by a single scale. Note: a Sans Serif font has been applied in the text of this dissertation in order to discern the observed variables from the latent variables.



**Figure 3.12: Core Structural Equation Model**

## COMMITMENT TO THE STUDY (PRE) AND COMMITMENT TO THE STUDY (POST)

Cummings (1997) in his study focusing on the influence of stakeholder involvement in evaluation studies on the use of the evaluation information, found increased stakeholder commitment to an evaluation to have a positive influence on the use of the evaluation information. Other authors concur with this finding (Alkin et al., 1979; Dickey, 1980; Hudson-Mabbs, 1993; Patton, 1997; Vlahov, 1989). With regards to this study, two latent variables focus on stakeholder commitment to the study, *Commitment to the Study (Pre)* and *Commitment to the Study (Post)*.

The items for the scales of each have been adopted from the work of Cummings (1997) and Hudson-Mabbs (1993) and focus on stakeholders' perceptions of the utility of the evaluation information, the need for the study, the appropriateness of the study timing and the extent to which stakeholders support the evaluation. The items of the two scales are similar to each other, except where changes in tense have necessitated some minor change in wording.

## PROCESS USE

Patton (1997, 1998) identified 'process use' as changes in the thinking or behaviour of those involved in an evaluation resulting from the learning that takes place in the evaluation process, as well as changes to program or organisational culture and processes. He considers his previous conceptualisation of evaluation utilisation prior to his identification of process use, to be too narrow (Patton, 2004). *Process Use* has been included as a predictor variable in the Core Model, with the expectation that it will be both influenced by other variables and will influence the outcome variable. It is anticipated that an increased process use will lead to increased influence of the evaluation's information, although it is possible that a high level of process use may pre-empt the influence of the final evaluation report (Hudson-Mabbs, 1993).

The observed *Process Use* variable is measured by one scale. The items of the scale focus on various types of influence identified by numerous authors in typologies of evaluation influence, such as Alkin et al. (1985), Cousins and Leithwood (1986, 1993), Hudson-Mabbs (1993), Leviton and Hughes (1981), Shulha and Cousins (1997). The scale items focus on the extent to which stakeholders' involvement in the evaluation process has influenced their thinking about the program, thinking about issues beyond the program, any decisions they were involved in making, or their practice.

## **OUTCOME MEASURES OF THE CORE MODEL TO BE TESTED**

The Core Model has been developed for testing four different areas of influence outcome measures: *Influence of Use of Program Theory in the Final Report (IUPTR)*, the *Influence of Involvement in Program Theory Elaboration (IIPTE)*, the *Use of Evaluation Information*, and the *Influence of Study Findings (ISF)* (see Figure 3.12). Each outcome variable considered below, is represented in the model by a single observed variable composed of one scale. Each of the outcomes was used separately as they are conceptually different. In addition, combining the outcome variables in the Core Model into a single outcome variable would have prevented the identification of the particular influence of the predictor variables on each of the separate outcomes. If only one structural model were developed, valuable outcome information might have been lost due to efforts to modify it to a form that would fit the data.

### **INFLUENCE OF STUDY FINDINGS (ISF)**

*Influence of Study Findings (ISF)* is the latent outcome variable of the first version of the Core Model developed for this study. This outcome measure is intended to focus on the more 'traditional' strains of evaluation influence detailed earlier, such as conceptual, instrumental and strategic (Alkin et al., 1985; Cousins and Leithwood, 1986, 1993; Cummings, 1997; Hudson-Mabbs, 1993; Leviton and Hughes, 1981; Patton, 1997; Weiss, 1972). However, the term 'influence' has been used purposefully in the wording of the scale items, and in the factor label, to indicate an expanded appreciation of evaluation

influence beyond the impact of the final report. Furthermore, the items of the scale measuring *Influence of Study Findings* request information regarding the evaluation and the evaluation's information, not simply the final report, reflecting the less restricted view of the impact of an evaluation's information; termed 'influence' by Kirkhart's (2000) reconceptualisation of evaluation use.

#### INFLUENCE OF USE OF PROGRAM THEORY IN THE FINAL REPORT (IUPTR)

*Influence of the Use of Program Theory in the Final Report (IUPTR)* is an important model variable. As with the outcome variable of Model 1, the first version of the core model, the term 'influence' has been used in both the wording of the scale items and in the label of the observed factor, again reflecting the impact of Kirkhart's (2000) expanded appreciation of evaluation use. The areas of influence that are the focus of the scale include influence on the stakeholder's thinking, practice, and decision-making. *Influence of the Use of Program Theory in the Final Report* is the outcome variable of the second versions of the Core Model developed for this study.

#### INFLUENCE OF INVOLVEMENT IN PROGRAM THEORY ELABORATION (IIPTE)

The effect of program theory on the use of an evaluation's information, is a primary focus of this study. Furthermore, the influence of involvement in any processes to elaborate or detail the theory underpinning the program is expected to have an impact on those involved. The literature suggests stakeholder involvement in any process to elaborate the program theory, or achieve consensus regarding program theory, may increase their awareness and understanding of the program, encourage reflection regarding their practice, lead to changes in their practice and inform decisions regarding the program evaluated (Cummings et al., 2001; Huebner, 2000; Milne, 1993). The items of the scale focus on evaluation impact on the thinking, practice and program decisions to be made by stakeholders. *Influence of Involvement in the Program Theory Elaboration (IIPTE)* is the outcome measure of the third version of the Structural Model developed for this study.

## USE OF EVALUATION INFORMATION- IMPORTANCE AND LIKELIHOOD

The *Use of Evaluation Information* latent outcome variable is represented by four observed variables, Importance of Use 1, Importance of Use 2, Likelihood of Use 1, Likelihood of Use 2. The scales of all four have been taken from an instrument included in the work of Cummings (1997). The 52 items of the scales (13 sets of four) focus on the importance (two from each set) and likelihood (two from each set) of a variety of possible uses of the evaluation's information. With regards to the items of these scales, each set of items focuses on specific uses of the evaluation information. *Use of Evaluation Information* is the outcome measure of the fourth version of the Core Model developed for this study.

In consideration of the two Importance of Use sub-scales, and the two Likelihood of Use sub-scales indicating the observed variable *Use of Evaluation Information*, a decision was made to combine the four to represent one latent variable labelled the *Use of Evaluation Information*. A reliability analysis of the scales ( $r_{ii} = 0.83$ ) indicated the decision to combine the four to be a reasonable one.

## SUMMARY

Eight versions of the Core Model are to be tested and comparisons of model fit considered, with a view to gaining insight into the influence of the predictor variables on the outcome variations. Each outcome variable was included in two versions of the Core Model, one including implementation theory and the other causative theory as the program theory predictor variable. This strategy was expected to facilitate model fit in view of the small sample size of the study. Furthermore, in consideration of the simpler models, it was expected that the influences of the predictor variables on the outcome variables, both direct and indirect, would be clearer. A more elaborated model which reflects more fully the components of the conceptual model and identified earlier in this chapter, was also tested. This elaborated model is presented in the following section.

## ELABORATION OF THE STRUCTURAL MODEL

Further elaboration and testing of the structural model was undertaken following the assessment of the model fit for the eight Core Model versions. An elaborated model includes all of the variables constructed, to reflect as closely as possible to the conceptual model. The additional five latent variables included in the elaborated structural model are: *Organisational Environment Characteristics*, *Stakeholder Characteristics*, *Evaluator Characteristics*, and *Study Involvement* measured both Pre and Post study. The five variables are briefly described below. A version of the intended final Elaborated Structural Model is provided in Figure 3.13

### THE ELABORATED STRUCTURAL MODEL VARIABLES

The literature relevant to each of the additional variables tested in the elaborated model variable is briefly reviewed below. A fuller consideration of pertinent literature was provided in Chapter II.

#### ORGANISATIONAL LEARNING ENVIRONMENT CHARACTERISTICS (OLE)

Numerous authors have argued that the environment of an organisation has a significant influence on the behaviour of its employees (e.g. Argyris, 1999; Bartlett and Ghoshal, 1995; Dilworth, 1995; Luthans, 1998; Preskill and Torres, 1999a). The work of Preskill and Torres (1999a, 1999b, 2000a) considers the importance of evaluative inquiry for organisational learning. They argue that for evaluative inquiry to be successful, an organisation must have the infrastructure to support it (1999a). This proposal is pertinent to the present study and therefore will be included in the elaborated version of the Core Model. The OLE characteristics factor is measured by five scales, Organisational Environment (OE), Leadership Practice (LP), Personal Practice (PP), Teamwork (T), and Use of Information (UI). A review of literature relevant to each scale is presented in Chapter IV, which details the interview schedule.



## EVALUATOR CHARACTERISTICS

Evaluators are generally considered to have an influence on the use of an evaluation's information (Alkin and Dailak, 1985; Cousins and Leithwood, 1986; Greene, 1988; Hammond, 1983; Hudson-Mabbs, 1993; Mathison, 1994; Seigel and Tuckel, 1985). Therefore the influence of the characteristics of the evaluator has been included in the elaboration of the Core Model. The *Evaluator Characteristics (EC)* latent variable is indicated by four scales, Evaluator Practice 1 (EP1), Evaluator Practice 2 (EP2), Evaluator Practice 3 (EP3), and Evaluator Perception of Stakeholder Involvement (EPSI). The logic underpinning the development of each scale is given in Chapter IV.

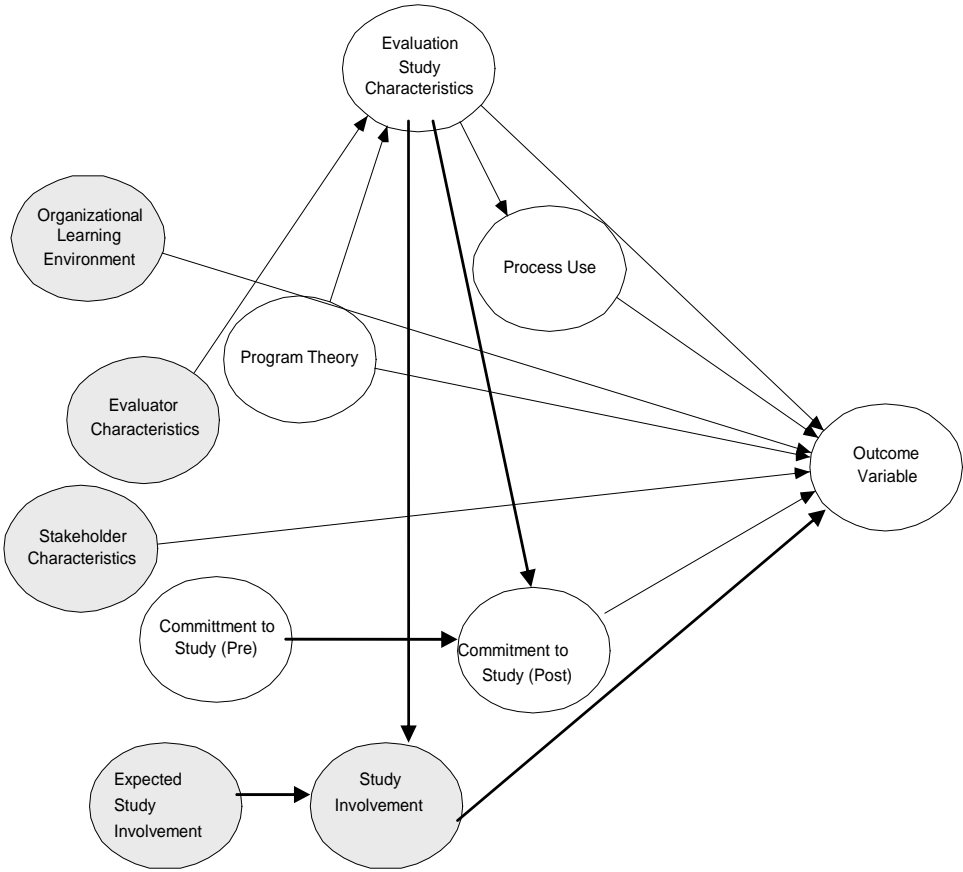
The characteristics of the stakeholders of an evaluation have been found by many authors to have an influence on the use of the evaluation's information (Cousins and Walker, 2000; Cummings, 1997; Hudson-Mabbs, 1993; Leviton and Hughes, 1981; Vlahov 1990). The characteristics of focus in this study are represented in the five observed variables indicating the latent factor *Stakeholder Characteristics*. They include Commitment to the Program, Education and Training, Opportunity to Use (Information), Commitment to the Organisation 1 and Commitment to the Organisation 2.

## STAKEHOLDER CHARACTERISTICS

### **Study Involvement**

A review of evaluation literature reveals a strong link between stakeholder involvement in an evaluation study and the use of its information (Cousins and Earl, 1995; Cummings, 1997; Patton, 1997; Preskill and Torres, 1997, 2000). Torres and Preskill (2000) suggest that when individuals come together to consider information, such as when stakeholders consider evaluation information, social learning occurs. Two latent variables focusing on study involvement have been included in the elaboration of the Core Model of the study. *Expected Study Involvement* focuses on the expectations of stakeholders regarding their involvement in the evaluation study process, while the *Study Involvement (Post)* variable focuses on the actual involvement of the

stakeholders in the evaluation process. The wording of the items of the two scales measuring the factors is the same, except where changes in the wording have been made to accommodate tense changes.



Note: Latent variables added in elaboration of Core Model have been shaded.

**Figure 3.13** Intended Version of the Elaborated Structural Model

## SUMMARY

The intent of the first portion of this section has been to detail the Core Model to be tested in this study. A description of the five latent predictor variables of the model and the four outcome variables has been included. As there are four outcome measures, eight versions of the Core Model have been developed for testing, one pair (one with the program implementation theory latent predictor variable and the other with the program causative theory latent predictor variable) for each outcome variable. The final portion of the section is focused on the further elaboration of the Core Model, with five latent predictor variables being included.

## CLASSIFICATION OF PROGRAM THEORY USE IN EVALUATION STUDIES

A review of published theory-based evaluation studies indicates that there is little consensus about what a program theory is. This lack of agreement means there is no widely accepted definition and hence about different degrees of theory in an evaluation study. Therefore, the first challenge in considering the use of theory-based evaluation information is to determine a way of classifying evaluations in a manner relevant to theory-based practice. Categorising evaluation studies in accordance with some key characteristics enables the comparison of similar theory-based evaluations.

In the Classification Matrix developed for this study and presented in this chapter the categories have been generated in line with four keenly debated areas of theory-based evaluation practice: the extent of use of two program sub-theories, causative theory and implementation theory; the degree to which a program's context is considered in an evaluation; the origins of the program theory used in an evaluation, i.e. social science theory, the stakeholders or the evaluators; and the level or complexity of theory use. An outline of the evaluation theory classification matrix is given in Table 3.1.

## CLASSIFICATION MATRIX

The Classification Matrix is a tool for categorising theory-based evaluation studies according to their use of program theory. Two sub-theories of program theory are considered: program implementation theory and program causative theory. These two sub-theories, distinguished by Weiss (1997b), were examined in the review of program theory literature undertaken in Chapter II, and earlier in this chapter. However, as they are very pertinent to the Classification Matrix, both are further considered in the section below.

### PROGRAM IMPLEMENTATION AND PROGRAM CAUSATIVE THEORY

Suchman (1967), in his book *Evaluative Research*, delineates two avenues for program failure. One is the failure to put intended program activities into operation as planned. The other is the program activities' lack of success at bringing about the intended outcomes. The first can be seen as the failure to carry out the program's implementation theory. The second can be seen as the failure of the causative theory.

A program's causative theory identifies the causal links between the program application, or intervening variables, and its intended effects. Identifying these intervening variables and carefully considering their effect in the implementation of the program, can provide policy-makers and program developers with useful information regarding a program's operation. Specifically, an explication of causative theory should identify a key mechanism or leverage factor upon which an intervention or treatment has been developed (Chen, 2003). Implementation theory, alternatively, considers how the program is conducted, and addresses the variables affecting the extent to which the program is implemented as planned (Scheirer, 1987; Weiss, 1997b).

Several authors (Bickman, 1987; Chen, 2004; Scheirer, 1987; Weiss, 1997a) have identified the ability to discern between the failure of a program's

causative theory and the failure of the program's implementation theory as a significant merit of theory-based evaluation. They argue that the delineation of program causative theory failure from program implementation failure is not possible unless it can be shown that a program was faithfully implemented, as intended, with reasonable levels of quality, intensity and fidelity. If a theory-based evaluation finds that a successful program has been implemented with fidelity to the program implementation theory, then the evaluation information lends support to the program's causative theory. Should the evaluation find, however, that the implementation of the program is not consistent with the implementation theory it is difficult, if not impossible, to attribute success or failure of the program (in terms of achievement of its intended outcomes) to the inappropriateness of either the implementation theory or the causative theory. As Bickman (1987) has found, there are numerous examples in the literature of interventions or programs found to have no effect and hence discredited, when in fact the intervention, as it was intended, was never implemented.

The investigation of both the implementation and causative theories of a program is a daunting task given the need to account for a large number of variables (Scriven, 1994; Weiss, 1997b). As Weiss (1997a) found, and a scan of published theory-based evaluations confirms, most theory-based evaluations focus on either the program's implementation theory or the program's causative theory. Furthermore, overwhelmingly theory-based evaluations consider the program's implementation theory.

One possible reason for this focus on program implementation theory rather than causative theory in theory-based evaluations, is that the measurement of variables mediating causative effects is a relatively recent endeavor which continues to face many challenges (Conrad and Buelow, 1990). Alternatively, it may be that evaluators are more familiar with the measurement of variables moderating program implementation. The measures adopted tend to be more practical and less abstract than measures of causative theory mediators. Weiss (1997b) identifies the valid measurement of mediating variables as a necessary development in the evolution of theory-based evaluations.

The Classification Matrix in Table 3.3 outlines a method of classification of the incorporation of both program implementation theory and program causative theory in a theory-based evaluation. The Classification Matrix categories and the classification process are considered more fully below. First, program implementation theory and program causative theory are delineated. The theory origin categories are then defined. The evaluation's use of each sub-theory is classified in terms of its origin (evaluator, stakeholder or social science theory), a consideration of context (no or negligible, minimal, substantial), and the complexity of the theory (no or negligible, minimal, substantial). Finally, a consideration of the relevance of program context to the program theory is undertaken. The remainder of the section focuses on the use of the Classification Matrix. An example of matrix use is then followed by the review and classification of 20 evaluation reports, the majority of which are theory-based. A summary of the evaluation report classification process follows, which includes some comparisons of theory-based evaluation practices and a discussion of obstacles encountered in the use of the Classification Matrix. The section ends with a determination of numerical measures based on the Classification Matrix data for each evaluation.

**Table 3.3: Matrix for Classification of Theory-Based Evaluations**

	Program Implementation Theory									Program Causative Theory								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
Program Evaluation Study	None or Negligible	Minimal	Substantial	None or Negligible	Minimal	Substantial	None or Negligible	Minimal	Substantial	None or Negligible	Minimal	Substantial	None or Negligible	Minimal	Substantial	None or Negligible	Minimal	Substantial

It may be the case that the evaluative undertaking has incorporated both the causative and the implementation theory of the program. Alternatively, the evaluation may have incorporated only the program causative theory or program implementation theory into the evaluation, or possibly neither. In summary, the Classification Matrix has been developed to classify a range of theory uses in theory-based evaluation.

#### Stakeholder Theory, Social Science Theory and Evaluator Theory

One of the more ardently debated issues pertaining to the practice of theory-based evaluation is the source of the theory (Chen, 1990b; Riggin, 1990; Trochim, 1985). Evaluation findings based on a social science theory may allow the prediction of long-term program outcomes, be linked back to current developments in social science theory, and contribute to social science knowledge, thus widening the use of the evaluation findings to areas beyond the program (Chen, 1990; Fitz-Gibbon and Morris, 1996; Weiss, 1997a). Alternatively, concerns regarding the use of social science theory include the concessions made to fit the program to an abstract social science theorem and the ability of stakeholders with no research background to grasp the esoteric meaning of such evaluation findings (Patton, 1989; Pawson and Tilley, 1996; Scriven, 1994).

Proponents of the use of stakeholders' theory in evaluation, however, point out that most programs originate from the tacit theories of stakeholders (Conrad and Miller, 1987; Pawson and Tilley, 1997; Wholey, 1987). Patton (1989) argues that the elucidation of the program theory from the stakeholders' perceptions and assumptions, is a practical inductive approach to program theory development that enables theory-based evaluation findings to be understandable and useful to key stakeholders. Consideration will most likely be given to the priority to be allocated to each stakeholder group in the construction of a program theory. There are, however, concerns that the use of stakeholders' theory to evaluate a program may limit the evaluation

somewhat, as it links the evaluation closely to the beliefs, values, and norms of the stakeholder group that may represent the problem the program is to address, and thereby introduces a potential source of bias (Shadish, 1987).

Both Social Science Theory and Stakeholder Theory have been included as theory origin categories in the Classification Matrix given here and where a particular social science theory is specified (e.g. Bandura's Social Learning Theory), it is noted. However, in many cases the social science theory referred to is deduced from a review of relevant conceptual and experimental literature and no explicit social science theory is identified.

Acknowledging the significance of the evaluator role in the program theory adopted in the undertaking of a theory-based evaluation, a classification of the evaluator's contribution to the Program Causative Theory and Program Implementation Theory is included as a category of the Classification Matrix. In considering descriptions of the theory-based approach offered in the relevant literature Mertens (1996, p. 354) writes,

...they described theory-based evaluation as an approach in which the evaluator constructs a model of how the program works, using models based on stakeholders' theory and/or available social science theory to guide question formation and data gathering.

There may be occasions where a theory-based evaluation is undertaken in accordance with a program theory that has been clearly explicated by program developers, faithfully followed by program practitioners, fully intact and ready for use, as is, by the evaluator endeavoring to undertake a theory-based evaluation. An examination of the relevant literature indicates, however, that these occasions are rare. More commonly, evaluators are likely to undertake one or more of the following courses of action: explicating the program stakeholders' implicit theories of program action; or choosing a pertinent social science theory; or generating a program theory themselves based on their experience and knowledge gained in interviews with those involved in



the program, document reviews, observations and knowledge gained from the academic literature.

Whichever strategy is adopted for making a the choice of the program theory upon which an evaluation will be based, an opportunity for the evaluator to influence the stated theory clearly is present. The values, training, social affiliations, academic and professional interests of the evaluator all have a potential to influence the program theory generated for an evaluation (Weiss, 1997a). Furthermore, the evaluator undertakes the evaluation with an awareness of the interests of the party contracting the evaluation and the purpose of the study.

The evaluator often has an opportunity to incorporate their perception of the program theory into the design of the evaluation. Their theory may not be explicitly stated, but it will often be seen in the variables considered as important in the determination of control groups, or the selection of particular demographic variables about which to collect data. The inclusion of these variables in the evaluation design is not incidental, but rather may be based on previous experience, discussions with stakeholders of the program under review, or the observation of program activities. The usefulness of the theory is often seen in the extent to which the variables targeted in the evaluation design provide insight into the evaluation findings.

#### NO OR NEGLIGIBLE USE, MINIMAL USE AND SUBSTANTIAL USE

Scriven, (1998) has presented a taxonomy of theories, which was adapted for use in the Classification Matrix in order to categorise the evaluation use of Social Science Theory, Stakeholder Theory and Evaluator Theory according to degree. The three levels of use are 'No or Negligible Theory Use', 'Minimal Theory Use', and, 'Substantial Theory Use'.

No or Negligible Theory Use may include some conceptualisations and accounts of program operation and effect, but no explication or testing of the conceptualisation is undertaken in the evaluation. The traditional “black box” evaluations fall into this category.

Minimal Theory Use is restricted to the explication and assessment of simple descriptive hypotheses regarding the program. The “basic two-step” described by Lipsey and Pollard (1989) as inquiry into one intervening mechanism between inputs and outcomes, easily falls into this category.

Substantial Theory Use involves the explication and assessment of an *a priori* theory incorporating program components, rationale, mediating causal links and expected outcomes. Lipsey et al. (1985) grouped 119 evaluation studies then divided them into non-theoretical, sub-theoretical and theoretical categories designed to discriminate between different levels of theoretical sophistication. They found that only nine per cent of the theory-based evaluations they examined explicated and tested an integrated program theory, while twenty per cent were non-theoretical evaluations.

## CONTEXT

Pawson and Tilley (1997) advocate the value of considering the context within which the program will be introduced, echoing the position of Stake (1967, 1991) and write, “In realist terms, it is the contextual conditioning of causal mechanisms which turns (or fails to turn) causal potential into causal outcome” (p. 69). Similarly, Chen (1990) expresses the importance of investigating contextual conditions to widen the understanding of causal inference. However, as highlighted by Pawson and Tilley (1997), program evaluation practice, in general, has failed to address the social conditions into which a program is introduced.

With regards to Implementation Theory, an evaluation’s consideration of the context within which a program operates may provide insight into the effect of these contextual variables on the program. Any adaptations made to

accommodate the effect of the contextual variables may, in turn, impact upon the causal mechanisms and outcomes of the program. Because of this, a comprehensive assessment of causative inference has been included in the matrix as a factor for classification in the consideration of theory origin, Evaluator Theory, Stakeholder Theory, or Social Science Theory.

#### NO OR NEGLIGIBLE USE, MINIMAL USE AND SUBSTANTIAL USE

Categories are presented for three levels of context usage. They are: 'no or negligible', 'minimal', and 'substantial'. Context is represented in the Matrix by the letter 'C' in the boxes classifying Stakeholder, Social Science and Evaluator Theory. The number attached to the C represents the level of Context use. No or Negligible Use (represented with a C1 notation) of contextual information refers to no consideration of program relevant contextual variables, or use only in a descriptive manner, as in an evaluation's description of a program. Minimal Use (represented with a C2 notation) of contextual information, includes a more detailed description of a program's context extending to simple conceptualisation hypotheses and an assessment of the effect of program context on the program's activities and practices (Program Implementation Theory), underlying intervening mechanisms (Program Causative Theory), and outcomes. Substantial Use (represented with a C3 notation) of contextual information, includes the integration of a program's contextual variables into any evaluation hypotheses stated regarding the program's causative theory or implementation theory and the expected outcomes.

#### HOW THE CLASSIFICATION MATRIX WORKS

Program Implementation Theory and Program Causative Theory are conceptual components of all programs, at some level. The Classification Matrix requires that both sub-theories be classified along three dimensions: i) the use of Stakeholder Theory, Social Science Theory, and Evaluator Theory, ii) the level or sophistication of theory complexity, and iii) the consideration of context in the theory used for the evaluation for both program

implementation theory and program causative theory. The application of this classification scheme to an evaluation study is given below.

#### AN EXAMPLE

The paper “Evaluating the Process and Outcome of a Garbage Reduction Program in Taiwan” reports a theory-based evaluation undertaken by Chen, Wang, and Lin (1997) of a program with the goal of reducing the volume of garbage output of a community in Taiwan. The article includes a figure outlining the causal processes of the garbage reduction program from “implementation to effectiveness” (p. 30). To empirically verify the program’s theory, including both causative and implementation theory, Chen et al. integrate a process evaluation and an outcome evaluation. Refer to Table 3.4 for the classification of this evaluation.

Considering the evaluation in terms of the Classification Matrix, it is noted that the evaluation design addresses and includes measures of both program implementation theory and causative theory. The program theory, both implementation and causative, was formulated through review of all relevant program documents and reports, and by means of interviews with the program director and designer. The article makes no mention of any relevant social science theory. Both the program implementation and program causative theory, as developed for the evaluation, are those of the stakeholder, explicated by the evaluator from significant consultation with the program designer and director, and a review of program documents.

The next step in classifying the evaluation for the Classification Matrix is the determination of the level of theory sophistication included in the evaluation design. With regards to implementation theory, a variety of methods were used to assess program implementation, to determine the extent to which the program was implemented as intended. These measures include site visits to the sanitation department to review government records and tours of the main streets of the targeted community to ensure public awareness of the program. Other methods used in the evaluation include measures of the daily volume of

garbage collected and the tracking of fines issued for garbage dumping. Measures of neighbouring community garbage volume were also tracked to ensure that any reduction in the garbage volume of the targeted community was not a result of dumping in the neighbouring communities. The variety and number of measures of implementation indicate substantial interpretation of the stakeholder program theory by the evaluator (Program Implementation Theory: Stakeholder Theory- Substantial; Evaluator Theory- Substantial).

With regard to Program Implementation Theory, Evaluator Theory has been incorporated to strengthen the study design. The evaluators have included measures in the evaluation design to examine implementation issues particular to the context within which the program has been introduced. The evaluator's consideration of these issues remained simple because the data collected to ascertain the effect of the program context on the program (as actualised), revealed the issues not to be a threat to faithful program implementation. Had the data revealed the issues to have had an effect on program implementation, the evaluators may have offered a more complex theory to account for the impact of these moderating issues on the program outcomes.

#### **CLASSIFICATION OF THEORY-BASED EVALUATION REPORTS**

Classifications of 21 theory-based evaluation reports in accordance with the Classification Matrix presented in the previous section, are presented below. The first seven studies (including the above example) are wholly or predominately program implementation theory-based. The eighth, ninth and tenth evaluation reports listed are based on both program implementation theory and program causative theory. The following seven reports emphasise the use of program causative theory. The final four evaluation reports included in the Classification Matrix are neither program implementation theory nor program causative theory-based, either because they have not intended to be, or because the theory use is inappropriate to support the classification of the study as 'theory-based'.

CLASSIFICATION OF EVALUATION REPORTS FOUND TO BE PRIMARILY  
IMPLEMENTATION THEORY-BASED EVALUATION REPORTS

**The Model-Guided Method for Monitoring Program Implementation**  
Brekke, J. (1987)

This paper presents a model-guided approach for monitoring program implementation. As an example, a model-guided monitoring evaluation of a community support program for the treatment of young adult schizophrenics in Madison, Wisconsin, based upon the Training in Community Living (TCL) model, is reported in the paper.

The model against which the implemented program is compared is taken from written descriptions of the TCL model, both in published literature and manuals (Program Implementation Theory: Social Science Theory and Stakeholder Theory- Substantial), and from conversations among program designers and managers of the program (Stakeholder Theory- Substantial). It is important to recognise that in this instance two levels of program stakeholders are represented: those responsible for the development of the TCL model; and those responsible for the implementation of that model in Madison, Wisconsin. The process by which the document review and stakeholder discussions were combined to develop the program model for the evaluation, is not clearly delineated. Therefore, it is difficult to assess the extent to which the evaluator has influenced the development of the model. Due to lack of evidence to the contrary, the evaluator theory contribution to the program theory upon which the evaluation is based, has been considered to be negligible (Program Implementation Theory: Evaluator Theory- No or Negligible).

The program's Causative Theory is not explicitly considered in this evaluation. The focus of the evaluation is clearly the program's implementation theory. The consideration of context in the model developed, however, is substantial, particularly with regard to population (schizophrenic disorder) and the support systems necessary to enable them to live and function successfully in the community. Program designers extended some

prescriptive elements and hypotheses of the model to include empirical specifications tied to the program context (Program Implementation Theory: Context C3). Refer to Table 3.4 for the classification of this evaluation.

### **The Application of Program Theory to the Evaluation of a Managed Mental Health Care System**

Leonard Bickman (1996)

The Fort Bragg Child and Adolescent Mental Health Demonstration program was developed and implemented to enable the consideration of the effect of a continuum of mental health services on treatment outcomes at the Fort Bragg military post and the cost of care per client, as compared to more traditional mental health services lacking the key feature of continuum of care. The evaluators developed the initial logic model of the program two months prior to program implementation and prior to interviews with staff, or a review of program documents. However, the final program theory model upon which the evaluation is based was elaborated through interview, document reviews and focus groups. It is clearly a substantial theory of intended program implementation (Program Implementation Theory: Evaluator and Stakeholder Theory- Substantial).

The evaluation does not address the causal theory underpinning the program (Program Causative Theory- No or Negligible). Additionally, as is stated in the report (p. 114), the evaluation does not directly address contextual factors, though sites selected to represent the comparison control sites were similar to the site of demonstration program implementation in terms of population size, demographic characteristics, geographic location, the availability of mental health services on and off post, military readiness and command structure. These contextual considerations with regard to program implementation are those of the evaluator, and are indicative of consideration of contextual variables with a potential to influence program implementation and the program outcomes (Program Context: Evaluator Theory- No or Negligible). Refer to Table 3.4 for classification of this evaluation report.

**The 1995-1996 Cleveland Community-Building Initiative Baseline Progress Report: Village Council Formation, Asset Appraisal, Agenda Formation, and Action Projects**

Milligan, S., Nario-Redmond, M., and Coulton, C. (1997)

This article reports on a 'baseline' evaluation undertaken of the Cleveland Community Building Initiative (CCBI). The study adopts a "theories of change" approach (p. 2) in which both program causative theory and program implementation theory are to be addressed. The evaluation reported in this instance was undertaken for a program under development, therefore, the four specific short-term outcomes measured in the evaluation address implementation issues. These short-term outcomes are relevant to longer term outcomes touching upon causative program theory, but causative theory is not considered in the evaluation, as reported.

The theory which guided the study was developed from a series of individual interviews and groups discussions with CCBI staff, board members and village council members. CCBI staff provided insight into the intended outcomes of the program and assisted in the determination of benchmarks used to measure the progress toward achievement of outcomes measured in this evaluation (Program Implementation Theory: Stakeholder Theory-Substantial/ Evaluator Theory- Substantial).

Although it is not explicitly stated in the report, it is understood that the context in which the program is implemented is well integrated into the program theory. The CCBI is a component of the work of the Cleveland Foundation Commission on Poverty which exists to address the problem of persistent poverty in the neighbourhoods of Cleveland. In turn, the intended theory which guides this study rests upon the ideas of CCBI staff, board, and council members. In this particular case, it is impossible to separate the program theory from the context of the circumstances into which it has been introduced (Stakeholder Theory Context- Substantial). The consideration of context in the Evaluator's contribution to the theory is suspected to be a great deal less endemic (Evaluator Theory Context- No or Negligible). No use of



social science theory is reported. Refer to Table 3.4 for classification of this evaluation.

**Effects of Program Implementation on Adolescent Drug Use Behaviour:  
The Midwestern Prevention Project (MPP)**

Pentz, M., Trebow, E., Hansen, W., MacKinnon, D., Dwyer, J., Johnson, C., Flay, B., Daniels, S. and Cormack, C. (1990)

The purpose of this study was to investigate the relationship between the implementation quality of school-based social influence programs and changes in the drug using behaviour of adolescents. In this example of theory-based evaluation, the causative theory upon which the implemented program rests is underpinned by social learning and problem behaviour theories of behaviour change, transaction and systems theories of environmental change and relevant communication theories. In this instance, the program's causative theory is considered to be adequately supported by previous research and assumed to be sound. It is presented in a manner descriptive of this program (Program Causative Theory: Evaluator Theory/ Social Science Theory/ Stakeholder Theory- No or Negligible Use).

In this case, program implementation was operationalised as 'adherence', 'exposure', and 'reinvention'. The theory underpinning the conceptualisation of program implementation and the measures operationalised have been developed by the program evaluators in light of a review of relevant literature (Program Implementation Theory: Stakeholder Theory/ Evaluator Theory- Substantial Use, Social Science Theory- Negligible Use).

Some of the authors of the evaluation are among those responsible for the development and implementation of the intervention, particularly the first author of the report, Mary Ann Pentz. As a result, the evaluation is reported in such a way that prevents separation of stakeholder theory and evaluator theory. With regard to the Classification Matrix of the present study, stakeholder theory and evaluator theory are considered to be the same.

The evaluation's attention to contextual conditions influential to program implementation and effect is considerable, particularly with regards to the target population's drug use previous to program implementation and demographic characteristics at baseline and at one year of program exposure (Context: Program Implementation Theory- Stakeholder/Evaluator/Social Science Theory- Substantial, Program Causative Theory- Social Science Theory/Stakeholder Theory/ Evaluator Theory- Negligible (Descriptive)- No or Minimal). Refer to Table 3.4 for classification of this evaluation.

### **A Parent-Targeted Intervention for Adolescent Substance Use Prevention**

Cohen, D., and Rice, J. (1995)

No relevant social science theory has been referred to in this report of an intervention developed to test the authors' hypothesis that parental influence plays a greater role than peer influence in the onset of adolescent drug use, although social science literature and previous research has been used to substantiate the theory. Specifically, the authors hypothesize that parents might be able to directly prevent their children from experimenting with illicit substances if they knew their children's friends and could limit contact with substance-using peers. In line with a quasi-experimental design, comparison and treatment groups were established and data collected through the administration of surveys to the parents and the students in both groups.

The program developers and the evaluators are one and the same in this instance. Therefore, the classification of the stakeholder and the evaluator contributions to the theory upon which the evaluation is based is the same. Study measures assess program implementation theory components, including intensity, participation/attrition, and attendance. The implementation theory of the program upon which the evaluation is based has been developed by the stakeholders/ evaluators (Program Implementation Theory: Stakeholder/ Evaluator- Substantial, Social Science Theory- Negligible).

The program theory targets parents of adolescents in grades five, six and seven. Therefore, the theory-based evaluation is empathetic to the specifics of this particular group. Previous relevant social science research is referenced to

highlight some of the implementation issues previously experienced with similar groups in similar programs (Context: Social Science Theory- Negligible). Data regarding demographic factors including gender distribution, ethnic distribution and parental marital status is reported, and study findings are considered with reference to pertinent demographic groupings. However, the influence of the demographics of the subjects involved in this particular implementation of the program is not considered in light of a stated theory (Context- Evaluator/ Stakeholder- Negligible).

Causal mechanisms linking parental influence and adolescent drug use are present in the description of the program, but these mechanisms are not assessed in the evaluation design (Program Causative Theory- Negligible). An explicit statement of the program's causative theory is not presented. The report describes the implementation of the program. The results of the study indicate the program failed to effect changes in the drug-taking of adolescents. The theory upon which the program is based is found to be unrealistic. Refer to Table 3.4 for classification of this evaluation.

### **Methodology for Evaluating Mental Health Case Management**

Bryant, D. and Bickman, L. (1996)

This article reports on a theory-guided evaluation of the implementation and quality of case management services in the Fort Bragg Child and Adolescent Mental Health Demonstration Project in Fayetteville, North Carolina. The evaluation design rests upon the comparison of the actual service provided with the ideal intended model of service provision. A multi-method, multi-informant approach, including questionnaires, naturalistic observation, chart reviews and interviews, is used to document a detailed description of case management intervention content.

The intended model of provision of case management services was obtained through a review of case management literature, program policies and interviews with three stakeholder groups (parents, clinic administration and case managers). Brainstorming sessions were then conducted with each group

for the purpose of sorting and rating issues pertinent to the program implementation model, with a view to the development of criteria for high-quality case management. The final synthesis of this information produced the program (Program Implementation Theory) logic model (Stakeholder Theory-Substantial). No social science theory was used in the development of the program implementation model guiding this evaluation (Social Science Theory- NM).

The evaluation component focusing on assessment of project quality tied the implementation theory of the project to the project's context, particularly the providers and administrators. This represents substantial use of program theory originating with the evaluator (Context: Evaluator - Substantial).

Furthermore, although not a component of the intended model of program implementation developed for the evaluation, in undertaking the evaluation, information was collected regarding the effect of work environment and culture of the organisation on the performance of case managers. This information is particularly contextually relevant (Context- Stakeholder - Substantial). No consideration of program causative theory is reported. Refer to Table 3.4 for classification of this evaluation. A checklist of quality indicators relevant to the program was taken from concept mapping exercises undertaken with providers, families and administrators to develop an 'enriched' view of good case management. Information about the program contract, the mission statement of the clinic and a review of pertinent literature was also considered in the checklist developed to tap into the "opinions" (p. 124) of families.

#### SUMMARY OF IMPLEMENTATION THEORY-BASED EVALUATION REPORT CLASSIFICATION

As mentioned earlier in this chapter, the first seven studies are wholly or predominantly program implementation theory-based. These evaluations were found not to rest upon a theory originating with social science theory, but instead theories originating with the stakeholder and the evaluator. In the case

of the implementation evaluation, the activities and outcomes specific to the evaluated program are the priority. Most often stakeholders are important information sources for the setting of standards by which the program is measured. Their insight into what was intended to be implemented comprises the model against which the evaluation findings regarding what was implemented, is compared. Of the implementation theory-based evaluation reports classified, only Brekke (1987) refers substantially to relevant social science theory, yet in this instance the social science theory guides the program monitoring approach of the evaluation, as opposed to the implementation theory of the program. Furthermore, implementation theory-based evaluations tend to take into account the context of the program more often at a complex level rather than a negligible level. This link to context is a reflection of the grounding of program implementation theory in the practices and processes of the program. The classification of each of the seven evaluation reports of this section is represented in Table 3.4.

CLASSIFICATION OF EVALUATION REPORTS FOUND TO BE BOTH CAUSATIVE  
AND IMPLEMENTATION THEORY-BASED EVALUATION REPORTS

**Northeast In-Home Care Evaluation Project**

Nutter, B., Dennis, D., Forth, C., Goa, D., Nutter, M., and Weiden, T. (1995)

The evaluation reported here is a very good example of a theory-based evaluation and is based upon a very comprehensive logic and structure model, incorporating intended program activities (Program Implementation Theory) and expected causative links (Causative Program Theory), of the Northeast In-Home Care Evaluation Project. The Northeast In-Home Care Project is an innovative child welfare program with a focus on the care and condition of children's lives through in-home positive parenting education delivered by in-home care providers.

The evaluators, who had previously worked with both the Northeast In-Home Care Program and the Community Assisted Parenting Program, developed the idealized logic model of the program after a review of relevant theoretical and evaluation literature and consultation with the stakeholders. Although a thorough description of the model development process is not included, it

seems evident that the evaluators developed the model, though stakeholders and social science literature were important information sources (Program Implementation Theory: Stakeholder Theory/ Evaluator Theory- Substantial, Program Causative Theory: Social Science Literature/Stakeholder Theory/ Evaluator- Substantial).

The model developed for evaluation use integrates the contextual conditions of the program, with particular attention to the fit of the program with the needs of the child and the parents (Context: Stakeholder/ Evaluator Theory- Substantial). Refer to Table 3.4 for classification of this evaluation.

### **Improving Access to Academic Support in Higher Education**

Webb, J., Zhang, C., and Sillitoe, J. (1999)

This article reports of an evaluation of a project undertaken by the Victoria University Student Learning Unit to develop “...stand alone in-context academic skills guides to assist students’ transition into university life” (p. 13). The evaluation reported is part of broader investigation by two of the authors into student learning in higher education.

The guides were distributed to those first year students enrolled at the Department of Applied Economics who considered themselves to be lacking the skills and information necessary for a higher education student, and in need of both information and guidance regarding the undertaking of their studies. The introduction of the guides for first year students rests upon the theory that students beginning their higher education experience tend to be more open to suggestion than the more experienced students with established study patterns. (Program Implementation Theory- Social Science Theory- Substantial).

**Table 3.4: Classification of Theory Use in Evaluation Reports found to be predominantly Implementation Theory-Based**

	Program Implementation Theory									Program Causative Theory								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
Program Evaluation Study	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial
Chen, H, Wang, J. and Lin, L (1997)	NM					C3			C3	NM					C1			C1
Brekke, J. (1987)			C3			C3*	C1			NM			NM			NM		
Bickman, L. (1996)	NM					C 1			C2	NM			NM			NM		
Milligan, S. et al. (1997)	NM					C3			C1	NM			C 1			C1		
Pentz, M., et al. (1990)	C 1					C3			C3	C1			C 1			C1		
Cohen, D., and Rice, J. (1995)	NM					C 1			C 1	NM			C 1			C1		
Bryant, D. and Bickman, L. (1996)	NM					C3			C3	NM			NM			NM		
Note:	<p>NM in a box indicates the theory origin was not mentioned in the evaluation report.</p> <p>C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based.</p> <p>The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).</p> <p>* Stakeholder theory, in his instance, accesses two levels of stakeholders, stakeholders at the larger program design level, and stakeholders responsible for local implementation of the program.</p>																	

The article traces the role of both social science theory and stakeholder theory in the development of the study guides. In consideration of the evaluation, it is impossible to differentiate between stakeholder theory and evaluator theory as the evaluators are key stakeholders. The evaluation is certainly based upon theory, both causative and implementation, supporting the development of the study guides. However, the main data gathering instrument; a questionnaire composed of items designed to elicit open-end responses from the students as to their thoughts, attitudes and insights regarding the guides, does not obviously appear to be linked to the underpinning theories (Causative Theory- Social Science/ Stakeholder/Evaluator- Substantial, Implementation Theory- Social Science/ Stakeholder/Evaluator- Substantial).

Insight, gained in a brief discussion by telephone with the first author of the report revealed that the questionnaire items were tied to the theories underpinning guide distribution (Implementation Theory) and expected impact in terms of shifts in the students' comfort with the higher education setting and culture (Causative Theory). An end-of-semester results analysis of student respondents was undertaken in an effort to determine the impact of the guide on the students' assessment outcomes. The findings of the evaluation are linked to the stakeholder/evaluator theory and social science theory underpinning the development and distribution of the guides and social science theory informing the development of the guides. The theories of the report, stakeholder/evaluator and social science, all pertain explicitly to the learning environment of a higher education setting (Context- Substantial). See Table 3.5 for classification of the evaluation report in the Classification Matrix.

### **An Ecological Assessment of Community-Based Interventions for Prevention and Health Promotion: Approaches to Measuring Community Coalitions**

Goodman, R., Wandersman, A., Chinman, M., Imm, P., and Morrissey, E. (1996)

The evaluation reported in this article is of a community coalition to prevent alcohol, tobacco and other drug abuse and related behaviour. The approach of the evaluation is "ecological" (p. 35), meaning that it includes assessment



measures that focus beyond individual behaviour to target multiple social levels, and follows the program along a continuum of stages relevant to community readiness. The conceptual basis underpinning this ecological approach is supported by relevant literature and research previously undertaken by the authors in the development of the approach.

The evaluation report focuses on three program stages: program formation; plan implementation; and, impact. The evaluation component that assesses program formation includes the development of one-page diagram models, by the evaluators (Program Implementation Theory- Evaluator Theory: Substantial) that are tied to problems to be addressed by the proposed intervention and the proposed activities as they appear in the grant application. Markers of faithful intervention implementation are then developed in collaboration with project participants (Program Implementation Theory: Stakeholder Theory- Substantial). Measures to collect data pertinent to the markers were developed by the evaluators in order to provide data regarding the fidelity of the project implemented to the intended implementation. This evaluative component monitors project formation and as such, reflects program implementation. The models developed address intended program implementation. Program causative theory is not referred to, nor is the use of relevant social science theory reported.

The evaluation component assessing 'plan implementation' focuses on the delivery of effective program and policy initiatives, as intended. In addition to monitoring project implementation, the data collection measures utilised by the evaluators were expected to encourage those responsible for implementing the program to think beyond implementation to intervention refinement and long-term community changes (Context: Substantial).

The third evaluation component (project impact) includes measures of community-level indicators such as "incidence of intake into treatment program, substance abuse-related seizures and arrests, *per capita* liquor sales and licenses issued, substance abuse-related deaths, rates of adolescent drug use and blood alcohol levels of pedestrians and drivers in fatal accidents".

Measures of mediating links between project implementation and project impact are also assessed, at two levels. Surveys were conducted with community key leaders. The questions asked were directed primarily at policy and organisational level. However, questions were also included to assess the key leaders commitment to the project and to determine the extent to which the project had increased the awareness of key leaders and their organisations regarding awareness and concern of relevant issues. Furthermore, a telephone survey of a random sample of adults was undertaken to assess changes in community attitudes and behaviours.

The theory underpinning the assessment of mediating links between project implementation and impact is, as reported, entirely based upon evaluator theory. Consideration of project context with regards to the theory underpinning both the program implementation theory and the program causative theory used in the evaluation is substantial (Causative Program Theory: Evaluator Theory- Substantial). The theory is based upon a careful consideration of program stages, program levels, and the problems facing community-based interventions (Context: Evaluator Theory- Substantial). Certainly, social science theory and literature used to substantiate the theory underpinning this program is particular to the provision and structure of community services (Context: Social Science Theory- Negligible (Descriptive)). Refer to Table 3.5 for classification of this evaluation.

#### SUMMARY OF CLASSIFICATION OF IMPLEMENTATION THEORY AND CAUSATIVE THEORY-BASED EVALUATION REPORTS

Three reports were found to be based on both program implementation and program causative theory. An evaluation based on a theory should offer a clear statement of the program theory (implementation, causative, or both) including intended outcomes. The development of measures to assess the program processes and activities should then be linked to the articulated program theory including program outcomes. The evaluation findings may then be easily tied back to the program theory. Alternatively, a theory-influenced evaluation might certainly consider theory supporting the program

and possibly even articulate a program theory. However, unless the information collected is relevant to the program theory in a predetermined manner, it is considered an evaluation influenced by theory, but not based on the program theory.

**Table 3.5: Classification of Theory Use in Improving Access to Academic Support in Higher Education Webb, J., Zhang, C., and Sillitoe, J. (1999)**

	Program Implementation Theory									Program Causative Theory								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
Program Evaluation Study	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial
Nutter, M. et al (1995)	NM					C3			C3			C1			C3			C3
Webb, J. et al. (1999)			C3			C3			C3			C3			C3			C3
Goodman, R, et al. (1996)	C1					C3			C3	C1			C1					C3
Note:	NM in a box indicates the theory origin was not mentioned in the evaluation report. C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based. The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).																	

## SUMMARY OF CLASSIFICATION OF IMPLEMENTATION THEORY AND CAUSATIVE THEORY-BASED EVALUATION REPORTS

Three reports were found to be based on both program implementation and program causative theory. An evaluation based on a theory should offer a clear statement of the program theory (implementation, causative, or both) including intended outcomes. The development of measures to assess the program processes and activities should then be linked to the articulated program theory including program outcomes. The evaluation findings may then be easily tied back to the program theory. Alternatively, a theory-influenced evaluation might certainly consider theory supporting the program and possibly even articulate a program theory. However, unless the information collected is relevant to the program theory in a predetermined manner, it is considered an evaluation influenced by theory, but not based on the program theory.

The evaluation reports authored by Nutter et al. (1995), Webb et al. (1999), and Goodman et al. (1996) are based on program theory including both implementation and causative theory. Nutter et al. (1995) and Goodman et al. (1996) are examples of comprehensive theory-based evaluations. In fact, the evaluation reported by Goodman et al. (1996) included three evaluations: a forecast evaluation, an implementation evaluation, and an impact evaluation.

The report by Webb et al. (1999) was selected not because it was stated to be theory-based, but because it was hoped the exercise of including a non-theory-based evaluation report would assist in illustrating the classification process used in formulating the Classification Matrix. By chance, the report turned out to be significantly theory-based. In a brief discussion with the first author of the report, Janis Webb, it became apparent that explication of the theory underpinning the development of the guides was a prerequisite for approval of intervention funding. The extension of the theory-base into the evaluation methodology and measurement development was difficult to determine by reading the report, due to the form of the primary data collection instrument items.

In other reports reviewed for the Classification Matrix, evaluation measures reflect pre-determined measures of program activities and processes. However, in the instance of the Webb et al. (1999) the questionnaire items were purposefully open-ended to enable students to respond freely and to accommodate the wide variety of responses that would surely come from a target population as diverse as first year university students. It was suspected that the evaluators were searching for information pertinent to the heavy theory-base of the intervention, but the purposefully generative nature of the data gathering instrument made this link difficult to specify. In the evaluation report the findings are linked to the program theory-base. A discussion with Janis Webb revealed that, in fact, there were pertinent items of information relevant to both intervention implementation theory and causative theory that were targeted in the data collected. The classification of each of the three evaluation reports included in the Classification Matrix as being both implementation and causative theory-based is represented in Table 3.6.

#### CLASSIFICATION OF EVALUATION REPORTS FOUND TO BE PRIMARILY CAUSATIVE THEORY-BASED EVALUATION REPORTS

##### **Collaborative Development of a Theory-Based Student Assessment for a Violence Prevention Program Evaluation**

Constantine, N. and Curry, K. (1998)

The theory-based evaluation of a five-year violence prevention demonstration program, involved a collaborative process for the development of a theory-based assessment instrument. A local evaluation work group was assembled for the purpose of eliciting the program logic model (Stakeholder Theory). The logic model of the work group was then related to a generic program theoretical framework that was a modified combination (Evaluator Theory) of two social behaviour theories; Fishbein's Theory of Reasoned Action and Fisher's Information/ Motivation/ Behavioural Skills Theory (Social Science Theory).

A list of program elements, originally offered by the work group in the development of the program logic model, were rated by the work group

(Stakeholder Theory) after they were organised (Evaluator Theory) in accordance with the generic social behaviour theoretical framework. These elements were representative of only the causative theory of the program. Measurement instrument items were developed in accordance with elements that fit the theoretical framework. There is substantial use of theory with regards to the assessment and measurement of the program's causative theory, elicited from stakeholders and related to social science theory (Program Causative Theory: Stakeholder Theory/ Evaluator Theory/Social Science Theory- Substantial). The actual program activities are not assessed in this evaluation, although they are presented in the original logic model developed by the stakeholder work group (Program Implementation Theory: Social Science/Stakeholder: No or Negligible).

Although the figures included from the report indicate that contextual environment factors were included in the original work group program logic model, context was not considered in the theories and models developed in the undertaking of the evaluation. Contextual conditions are implicitly considered in the evaluation design and carried through most strongly in stakeholder input and may enlighten some of the findings regarding causative variables mediating program activities and short or longer term outcomes, but the assessment of these variables is not part of the evaluation design (Context: Stakeholder Theory- Negligible). Refer to Table 3.6 for classification of this evaluation.

### **An Impact Evaluation of Project SNAPP: an AIDS and Pregnancy Prevention Middle School Program**

Kirby, D., Korpi, M., Adivi, C. and Weissman, J. (1997)

The curriculum for Project SNAPP, an AIDS and pregnancy prevention middle school program, is heavily theory-based, adopting both social learning theory and the health belief model. The evaluation of the project, reported in this article, is also theory-based adopting both social learning theory and the health belief model to determine how SNAPP affected mediating variables that might influence behaviour. Variables measured included knowledge, beliefs, attitudes and self-efficacy of treatment and control group students. The

evaluation clearly addresses the program causative theory in a substantial manner.

The project was designed and implemented in accordance with social science theories chosen by the authors to provide significant positive behaviour impact. Thus, the authors of the evaluation were also the project designers and responsible for implementation. The input of other stakeholder groups to project design or the project evaluation is not mentioned in the article (Program Causative Theory: Social Science Theory/ Stakeholder Theory/ Evaluator Theory- Substantial).

The effectiveness of peer educators in the delivery of programs to young people; is a program implementation theory present in social science literature, has been applied by the program designers to this project. The authors personally selected and trained the peer facilitators and observed peer educators each time the curriculum were implemented. They believe the project was implemented with fidelity. However, beyond the assurance of the authors that the program was implemented as intended, no measures of program implementation were incorporated into the evaluation design. Program implementation is presented descriptively in the evaluation report. Although peer educators were considered to be effective in program delivery, no information was gathered on how they impacted on program delivery (Program Implementation Theory: Stakeholder Theory- No or Negligible).

The influence of program context was considered in program delivery, particularly in the comparison of background characteristics of the treatment and control groups and baseline assessment. However, contextual conditions influential to both program causative theory and program implementation theory were only considered descriptively in the evaluation, as reported (Context- No or Negligible). Refer to Table 3.6 for classification of this evaluation.

### **Experimental Evaluation of a Modular Approach to Mobilizing Antitobacco Influences of Peers and Parents**

Biglan, A., Ary, D., Yudelson, H., Duncan, T., Hood, D., James, L., Koehn, V., Wright, Z., Black, C., Levings, D., Smith, S. and Gaiser, E. (1996)

This article reports an evaluation of an intervention to increase the effects of a school-based program to prevent tobacco use by adolescents. The theory-guided evaluation uses an 'experimental' time series design, in which survey results of member communities exposed to one, two, or none of the intervention components are compared to determine intervention impact.

In this instance, the evaluation authors were also responsible for the development and implementation of the program. As a result, the stakeholder theory and evaluator theory classification in the Classification Matrix are the same. As a direct result of the authors' dual-role involvement with the intervention, significant detail is reported regarding the theory underpinning and guiding the intervention, conception and development of the program, the theory supporting the chosen methods of program implementation (Stakeholder and Social Science Theory), and assumed causative links (Social Science Theory).

However, in the methods of the evaluation the causative theory of the program is predominantly considered with regards to the intervention's influence on relevant parent and adolescent tobacco-related knowledge, attitude and intentions (Program Causative Theory: Social Science Theory/ Stakeholder Theory/ Social Science Theory- Substantial). Only two measures were included to assess intervention implementation. Responses to these two measures were so highly correlated that only measures for the first item were reported. The item required the respondent to indicate whether or not they had seen or heard anything about preventing tobacco use among adolescents in the previous two months in each of 14 media types (e.g. radio, pamphlets, bumper stickers, posters, etc.). This assessment of intervention implementation is classified as Minimal in the Classification Matrix of this paper, and rests upon stakeholder/ evaluator theory (Program Implementation Theory: Evaluator/ Stakeholder Theory- Minimal Theory Use).



Intervention context is reported descriptively, in consideration of both intervention implementation and evaluation findings. Context is not an integral component of the theory guiding the evaluation with regard to either program implementation theory or program causative theory (Context: No or Negligible). Refer to Table 3.6 for classification of this evaluation.

### **Breaking the Silence about Sexual Abuse of Deaf Youth**

Mertens, D. (1996)

The evaluator brings to this theory-based “proactive evaluation” (Owen and Rogers, 1999, p. 171) an appreciation (Evaluator Theory) of emancipatory theory (Social Science Theory) to explicate and frame the implicit theories of administration, staff, parents and students of a residential school for the deaf, regarding the conditions at the school that permitted sexual abuse to occur (Implementation Theory), why the abuse occurred (Causative Theory) and what steps could be taken to prevent such abuse from occurring in the future (Implementation Theory/ Causative Theory). Although a causative theory resting upon social science theory and evaluator theory is carried through to the methodology of the evaluation, as reported; an implementation theory is not (Program Causative Theory: Social Science Theory/ Evaluator Theory-Substantial). Program implementation is presented in a descriptive manner to highlight the vulnerability of deaf children to sexual abuse (Program Implementation Theory: Social Science Theory/Evaluator Theory-Negligible).

Observations of school activities are included in the report results, and the effect of these activities on the occurrence of sexual abuse considered in congruence with information collected regarding the attitudes and beliefs of staff and administrators regarding why the abuse had occurred (Causative Theory). A consideration of prevailing power inequities which work prevail to maintain the silence among those with the least power, is central to emancipatory theory. Furthermore, sensitivity to the context (of the school) and the culture (deaf culture) underpins the entire report (Context: Social

Science Theory/ Evaluator Theory- Substantial). Refer to Table 3.6 for classification of this evaluation.

**Evaluation of the Exposure Effects of a Theory-Based Street Outreach HIV Intervention on African-American Drug Users**

Collins, C., Kohler, R., DiClemente, M. and Wang, M. (1999)

This article reports the implementation and evaluation of the Street Outreach to Drug Abusers- Community AIDS Prevention (SODA-CAP) Project in Birmingham, Alabama. The intervention was developed in line with social cognitive theory and the trans-theoretical model of change. The process evaluation of the project was undertaken concurrent with intervention delivery. Intervention outreach workers collected baseline data prior to implementing the outreach activities and also collected the evaluation data in five consecutive years (1997 and 1998).

The logic of the causative theory upon which the intervention is based is the stakeholders' adaptation of the social cognitive theory and the trans-theoretical model of behaviour change to the circumstances particular to this community AIDS prevention project. These social science theories clearly underpin the evaluation design in terms of the targeted information and information collection strategies adopted. Stages of change leading to the intended target behaviour were identified and operationalised. In this evaluation those responsible for the development and implementation of the intervention are also the evaluators (Program Causative Theory: Stakeholder/Evaluator/Social Science Theory- Substantial).

The project implemented was uncomplicated. Risk reduction behaviours were framed in role model stories and distributed with male condoms and lubricants. The role model stories were tailored to age, gender, stage of change, and target behaviour. The intervention implementation is presented descriptively. No assessment of program implementation is presented, other than rate of respondent exposure to intervention activities. Intervention

activities are not assessed in terms of intensity, duration or impact (Program Implementation Theory: Stakeholder/Evaluator- Minimal).

The context into which the program has been introduced is clearly considered in the development of the intervention's causative theory and in the quasi-experimental design of the study considering two intervention and two comparison communities in a repeated, cross-sectional sampling method. Furthermore, in the discussion section of the report, contextual conditions identified as influential in the intervention delivery and in the evaluation are considered (Context- Program Implementation Theory and Program Causative Theory- Social Science Theory/ Stakeholder Theory/ Evaluator Theory- Maximum). Refer to Table 3.6 for classification of this evaluation.

### **Exploring the Link between Service Quality and Outcomes: Parent's Assessment of Family Support Programs**

Herman, S. (1997)

The assessment of family support services provided by the State of Michigan's public mental health system reported in this instance is underpinned by a theoretical framework that delineates various aspects of quality as: accessibility, continuity, technical management and interpersonal interactions. However, although social science theory is the base of this theoretical framework, the adoption of this social science theory to the purpose of this study is the conceptualisation of the evaluator.

The questions developed for this assessment were based on parent-generated ideas regarding what constitutes program quality from a previous study undertaken by the author. These questions were organised into scales, linked by the author to the aspects of quality service provision as represented by the chosen social science theory. Although the social science theory has played a useful organising role, the theory upon which this study is based is the conception of the evaluator.

The study focus is on parents' perceptions of service provision quality and on the outcome measures of parental stress, parents' perceptions of their resources, parents' perceptions of program effectiveness and satisfaction with

services. These are all measures linked to underlying causative mechanisms represented by parents' attitudes, opinions, values and perceptions. Very little information is given regarding the actual implementation of services. The focus of this assessment is clearly on a particular causative theory of quality service provision (Program Implementation Theory: No or Negligible, Program Causative Theory: Evaluator Theory- Complex; Social Science Theory- Minimal).

In the design of the evaluation, considerable information is gathered regarding the characteristics of the families included in the study. Although the evaluator hypothesis generating the targeting of these bits of information is not made explicit in the evaluation, it must be assumed that the evaluator had a reason to choose to collect this data and consider the influence of these variables (Context: Program Implementation Theory- Evaluator Theory- Minimal).

The report gives negligible consideration of the contextual conditions influential to service provision. Because service implementation is not a focus of the article, details of the context within which services are delivered are not considered here, and their potential influence does not play a role in the theory upon which the assessment is based (Context: Causative Theory- No or Negligible). Refer to Table 3.6 for classification of this evaluation.

### **Short-term Substance Abuse Prevention in Jail: A Cognitive Behavioural Approach**

Peyrot, M., Yen, S. and Baldassano, C. (1994)

The evaluation reported in this instance is of a cognitive-behavioural program delivered in the Baltimore City Jail, emphasising social skills training for substance abusers. The evaluators, who are also the program designers, have adopted the theory underpinning the program from a review of relevant social science literature. The program causative theory is the predominant theory upon which this evaluation is based. Evaluation instruments were designed to assess the impact of the program on client knowledge and attitudes in six content areas (Program Causative Theory: Stakeholder Theory- Substantial/ Evaluator Theory- Substantial).

Although no clear model of program implementation is delineated, it is clearly described and process measures relevant to program implementation are included in the evaluation design. These measures include program completion rates, group leader ratings of client participation and client ratings of group leader performance and skills. The data collected allows insight into program operation particularly when considered in light of findings regarding the program's causative theory. However, a number of program implementation issues were not assessed, enabling a variety of finding interpretations (Program Implementation Theory: Stakeholder Theory-Minimal/ Evaluator Theory- Substantial).

The context of the program is considered in the assessment of client background (demographics and criminal involvement) and in the assessment of client knowledge and attitudes (Program Causative Theory Context-Substantial). The influence of the jail setting on the delivery of the program is considered, yet no data is gathered to the hypotheses (Program Implementation Theory Context- No or Negligible). Refer to Table 3.6 for classification of this evaluation.

#### SUMMARY OF CAUSATIVE THEORY-BASED EVALUATION REPORT CLASSIFICATION

Causative theory-based evaluations often tend to originate with social science theory. The mechanisms by which a program's impact is mediated are often recognised as shifts in attitude, values and/ or beliefs. Theories linking acquisition of knowledge, attitude shifts and behaviour changes represent the grander theory of human society within which programs and interventions are implemented. The causative theory adopted by most theory-based evaluations considers human behaviour in the context of the program to be evaluated. The mediating variables tend to be abstract and difficult to operationalise with a any good degree of reliability and validity.

The use of two different levels of social science theory was brought to light by the process of social science theory classification for the Classification Matrix.

In instances where a grander social science theory has been referenced in a particular report, it has been noted in the text. Often however, the social science theory referred to has come to the evaluation by way of either evaluator or stakeholder review of relevant social science literature; conceptual and/or experimental. Possibly, in most cases, the theory developed based on a review of social science literature could be linked to a grander, social science theory, but the step has not been taken in the evaluation. In the case of Peyrot et al. (1994), the social science review was particular to substance abuse and social skills training, helping to identify contextual conditions particular to the problem targeted by the intervention. In this instance, the evaluators were content to not tie the program theory to a grander social science theory.

A look at the Classification Matrix shows the seven evaluations emphasising program causative theory, in all cases but one, originate with use of social science theory in conjunction with the use of stakeholder and evaluator theory. Substantial consideration, however, of the contextual conditions of the program have been undertaken in less than half of the studies. The classification of each of the seven evaluation reports of this section is represented in Table 3.6.

#### CLASSIFICATION OF EVALUATION REPORTS FOUND TO BE NEITHER CAUSATIVE NOR IMPLEMENTATION THEORY-BASED EVALUATION REPORTS

##### **Theory-Based Evaluation of an Experimental Program**

Dagenais, F. (1978)

The theory-based evaluation reported in this article adopts a quasi-experimental approach, and evaluates an experimental teaching program which utilises an effective teaching model. Three theoretical hypotheses regarding program implementation drive this particular evaluation. However, the basis of these three hypotheses is the extent to which the experimental classrooms and the comparison classrooms differ in the occurrence of 28 variables linked to the effective model of the experimental teaching program.

**Table 3.6: Classification of Theory Use in Evaluation Reports found to be Primarily Causative Theory-Based**

	Program Implementation Theory									Program Causative Theory								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
Program Evaluation Study	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial
Constantine, N. and Curry, K. (1998)	NM			C1			NM					C1			C1			C1
Kirby, D., et al. (1997)	NM			C1			C1					C1			C1			C1
Biglan, A., et al. (1996)	NM				C1			C1				C1			C1			C1
Mertens, D. (1997)	C1			NM			C1					C3	NM					C3
Collins, C., et al. (1999)	C1				C3			C3				C3			C3			C3
Herman, S. (1997)	C1			NM			C2				C1		NM					C1
Peyrot, M., et al. (1994)	C1				C1				C1	C1					C3			C3
<b>Note:</b> NM in a box indicates the theory origin was not mentioned in the evaluation report. C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based. The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).																		

The study data was gathered by the Stanford Research Institute (SRI) utilising the Classroom Observation Instrument, which involves five minute observation periods. Although 120 variables were recorded as having occurred in the five minute observation periods, only 28 were selected by the SRI as relevant to the particular model of the teaching program. These 28 variables were then classified by five program stakeholders responsible for program planning, development and implementation (Stakeholder Theory) of the experimental teaching program. Context of the program is represented in the 28 variables specified in the three evaluation hypotheses. These variables are taken from observed classroom activities and the extent to which they are relevant to the program theory determined by stakeholders. No mention is made of contextual variables.

The fit of actualised experimental program implementation to intended implementation is not considered. The evaluation acknowledges the theory of the experimental teaching program, though a model or theory of program implementation is not represented in the report. Certainly, the variables chosen for analysis are conceptually linked back to the program theory, but the basis of the analysis is simply the number of times these activities occur. In fact, only the evaluator's theory is acknowledged as it must be assumed from the report that the hypotheses tested in the study are those of the evaluator. The premise of the evaluation hypotheses fit best into the Minimal category, and then only for the program evaluator's contribution to the Program Implementation Theory used for the evaluation. The causative theory of the program is represented in the 28 variables used for program comparison, and therefore is not considered in the undertaking of this evaluation.

This evaluation report is particularly interesting as it is often referenced as the first theory-based evaluation report. Although the definition of a theory-based/ theory-driven evaluation presented in this report agrees with other definitions, the actual evaluation design and practice reported do not reflect what has been defined. It presents a confusing picture of theory-based evaluation. Refer to Table 3.7 for classification of this evaluation.



This report was provided in a response to a request sent to David Grembowski for theory-based evaluation reports. Prior to sending the report reference, Mr Grembowski requested a brief definition of what was meant to represent a theory-based or theory-driven evaluation.

Reading the report, however, revealed it to be substantially theory-influenced, but not theory-based. The goal of the intervention evaluated was to raise the level of self-efficacy of older adults for engaging in health promoting behaviours and maintaining health through the adoption of beneficial changes in health behaviour. The theory that self-efficacy - that an individual can influence their health through modified behaviour - as presented in the paper reported here, is central to Bandura's social learning theory and enlightening in the explanation of health behaviour change and maintenance. This theory has been adopted from relevant social science theory by the evaluators. Stakeholders, other than the evaluators (who are also responsible for intervention developments) are not mentioned.

### **Cost and Outcomes of Medicare Reimbursement for HMO Preventative Services**

Patrick, D., Grembowski, D., Durham, M., and Beresford, S. (1999)

The evaluators in a quasi-experimental design introduced the intervention with a control group for comparison. No intervention-specific models were presented. No *a priori* intervention implementation model was developed, nor were measures of services delivered linked to a choate logic of intended implementation (Program Implementation Theory: No or Negligible).

Although the theory of self-efficacy was described early in the article, no measures are undertaken of mediating causal links between preventive services and the intended goals of a reduced rate of hospitalisation, the use of non-preventive ambulatory care services, the total cost of care, or the rate of expected health-status decline among treatment participants (Program Causative Theory: Evaluator/ Social Science Theory- No or Negligible Use).

Although contextual conditions relevant to the intervention and the evaluation findings are included in the discussion section of the report, they are incorporated only to explain review findings. Contextual factors are not considered in the investigation as part of the program implementation or causative theory (Context: Evaluator/ Social Science Theory- No or Negligible). See Table 3.7 for classification of the evaluation report in the Classification Matrix.

**Table 3.7: Classification of Theory Use in Evaluation Reports found to be Neither Causative or Implementation Program Theory-Based**

	Program Implementation Theory									Program Causative Theory								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
Program Evaluation Study	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial
Dagenais, F. (1978)	C1			C1			C1			C1			C1			C1		
Patrick, D., et al. (1999)	C1			NM			C1			C1			NM			C1		
Sussman, S., et al. (1997)	C1			C1			C1			NM			NM			NM		
Sandler, I., et al. (1992)	C3				C3			C3		C1			C3			C3		
<b>Note:</b> NM in a box indicates the theory origin was not mentioned in the evaluation report. C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based. The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).																		

### **Implementation and Process Evaluation of a Student “School-As-Community” Group**

Sussman, S., Galaif, E., Newman, T., Hennesy, M., Pentz, M., Dent, C., Stacy, A., Moss, M., Craig, S. and Simon, T. (1997)

A component of a drug abuse prevention program implemented in continuation high schools, is the focus of this evaluation. Continuation high schools provide a supportive educational context for students who for emotional, behavioural or functional reasons, such as substance abuse, are not able to remain in California’s comprehensive school system. The ‘school-as-community’ component was developed by some of the authors of this evaluation as a part of a school-based drug abuse prevention program in which targeted high risk groups are engaged in a process of change from within the group.

The evaluation focuses on the program implementation and process. Implementation quality is referred to as the similarity of the program actually implemented, to the program implementation as intended. The process evaluation investigates the perceived quality of the program in terms of enjoyability, topic relevance and perceived helpfulness. The premise of this component of the investigation is that a well-received program is more likely to have a longer-term impact on behaviour. This theory has been developed by the authors through a review of relevant social science literature authored both by themselves and by others.

Possibly due to the dual role of at least some of the evaluation authors as program developers, the implementation theory underpinning the program is justified with a significant use of relevant literature and theory, yet no logical theory of program implementation is presented. The measures of program implementation adopted indicate the program was implemented, yet are not framed well in terms of how the intended program was to be implemented. Few measures of duration, quality or intensity of program-related events are included. The theory underpinning the program is descriptively well presented, but methods and design of the evaluation are not linked clearly back to the stated theoretical basis, nor is insight given into possible

benchmarks that might determine an acceptable level of program implementation (Program Implementation Theory: Stakeholder Theory/ Social Science Theory- No or Negligible).

Negligible consideration is given to the context into which the program has been delivered. For example, it is not known what other types of events were operating at the school at the time the 'school-as-community' program was implemented. The influence of contextual issues is considered in the conclusions of the study, yet has not been addressed in the consideration of intended program implementation, or in the design of the evaluation. Information regarding the demographic details of the participant students was collected in the process of the evaluation. However, the information was not used in any way to provide insight into evaluation findings, or to help to explain the outcomes of the program (Context- No or Negligible).

No causative theory has been considered in this evaluation. It is clearly an implementation/ process evaluation. See Table 3.7 for classification of the evaluation report in the Classification Matrix.

### **Linking Empirically Based Theory and Evaluation: The Family Bereavement Program**

Sandler, I., West, S., Baca, L., Pillow, D., Gersten, J., Rogosch, F., Virdin, L., Beals, J., Reynolds, K., Kallgren, C., Tein, J., Kriege, G., Cole, E. and Ramirez, R. (1992)

This article reports an evaluation of an experimental program for family bereavement program. The evaluators of the theory-based intervention are also the program developers, resulting in the singular representation of stakeholder and evaluator (Stakeholder and Evaluator Theory) contribution to the program theory upon which the evaluation is based. The program theory or model was developed by the stakeholders based on a review of pertinent social science literature (Social Science Theory) and specifies mediating causal links from parental death to mental health problems in children (Program Causative Theory). The two major phases of the program; a family grief workshop and a family advisor program, contained components with clearly hypothesised

causal links to psychological symptomatology. Instruments appropriate to the nature of each selected mediating variable were adopted for data collection.

Classification of this evaluation report has presented some particular difficulties. The mediators the evaluators chose to include in their model of the program are, in fact, moderators. The causative theory they have explicated is inappropriate. Therefore, the evaluation's use of causative theory has been classified as negligible, even though considerable theoretical work was undertaken (Program Causative Theory: Social Science/ Stakeholder/ Evaluator Theory- No or Negligible).

A process evaluation, which considers the occurrence of proposed program activities was also reported. The questionnaire, parent checklists and both family advisor checklists and notes were reviewed to ensure the intended program was implemented as faithfully as possible. No theory is clearly explicated which encompasses program implementation. No theories are extended regarding variables moderating program implementation (Program Implementation Theory: Stakeholder/Evaluator Theory- Minimal).

Although the program theory investigated in this evaluation has been conceptualised with regards to a specific population, namely children who have experienced the death of a parent, specifications pertinent to the local implementation of the program are not discussed. However, a considerable amount of information regarding participant demographic characteristics has been collected to enable insight into the influence of the program activities (Context: Program Causative Theory- Substantial). See Table 3.7 for classification of the evaluation report in the Classification Matrix.

#### SUMMARY OF CLASSIFICATION OF EVALUATION REPORTS FOUND TO BE NEITHER CAUSATIVE NOR IMPLEMENTATION THEORY-BASED EVALUATION REPORTS

In the selection of evaluation reports, a reasonable effort was put into locating theory-based evaluations incorporating program causative theory in an appropriate manner. The experience, however, confirms the findings of Weiss (1997a) that program theory-based evaluations tend to be based on

implementation theories rather than causative theory. The task was made more difficult by the inappropriate articulation of program causative theory (Sandler et al. 1992), the lack of causative theory articulation (Dagenais 1978), or confusion regarding theory-based evaluation practice and theory-influenced evaluation practice (Patrick et al. 1999, Sussman et al. 1997). Therefore, three of the four studies of this section were not at first selected to fit this category.

## **SUMMARY OF CLASSIFICATION PROCESS**

The 21 evaluation reports reviewed and classified above are all represented in Table 3.8. This classification exercise has provided useful insight into some of the points of confusion facing the undertaking of a sound theory-based evaluation. A few of the numerous sources of variability contributing to the diversity evident in theory-based evaluation practice, have also been highlighted. These include: the level of contribution of social science, stakeholders and evaluators to the elaboration of program implementation theory and program causative theory, the level of theory included in the evaluation process; and the consideration of the contextual conditions into which the program has been introduced. Nonetheless, this undertaking has collected a reasonable sample of what is presently accepted as theory-based evaluations, noted good theory-based evaluation practices and raised some questions regarding other theory-based evaluation practices. The introduction of a numerical measure to the program theory classification process in the final section of the chapter enables further analysis of the data. The concept of data analysis of the program theory use classification is considered again in Chapter VI of this work. Points of interest that arose in the categorisation process are discussed below.

## **EVALUATOR THEORY**

In the conceptual processes leading to the development of the Classification Matrix, it was understood that awareness that the theory of the evaluation and the theory of the program were separate. However, with the inclusion of Evaluator Theory category as a theory origin, it was acknowledged that in

many instances the contribution of the evaluator to the program theory is incorporated into the evaluation approach and design (e.g. Bickman, 1996), even though the evaluator contribution to the program theory upon which the evaluation is based may not be clearly delineated in the evaluation report. For example, the evaluator may advocate a particular evaluation approach that carries into the evaluation a social science theory they believe to be useful in the context of the program. Consider, for example, Mertens' (1997) choice to undertake a proactive evaluation with an emancipatory approach.

In short, the inclusion of Evaluator Theory as a category of program theory origin has proved a particularly useful means of flagging the contribution of the evaluator to the program theory in ways distinct from the program theory or model explication process. It is suspected that interviews with evaluators might reveal further specific evaluator theory contributions than have been recognised in this review of reports, and allow insight into the reasoning behind the tactics implemented in the process of the evaluation. Certainly, the evaluator may have considerable influence over which variables are assessed with regard to the program's causative theory, but the evaluator's theory regarding program implementation and the inclusion of measures of contextual variables, with a potentially moderating influence in the evaluation design, is at times a less obvious theory contribution. The routine collection of demographic information may provide valuable insight into program operation and evaluation findings (Bickman 1996). The selection of particular variables over others may be indicative of program theory hypotheses held by the evaluator. Due to the often implicit nature of the evaluator's contribution to the program theory version upon which the evaluation is based, however, clear identification of the evaluator component is difficult.

**Table 3.8: Classification Matrix Table with Classifications of all Twenty-one Studies.**

	Program Implementation Theory									Program Causative Theory								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
Program Evaluation Study	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial
Chen, H,Wang, J.and Lin, L (1997)	NM					C3			C3	NM				C1		C1		
Brekke, J. (1987)			C3			*C3	C1			NM			NM			NM		
Bickman, L. (1996)	NM					C 1			C2	NM			NM			NM		
Milligan, S., et al. (1997)	NM					C3			C1	NM			C 1			C1		
Pentz, M.,et al. (1990)	C 1					C3			C3	C1			C 1			C1		
Cohen, D.,and Rice, J. (1995)	NM					C 1			C 1	NM			C 1			C1		
Bryant, D.and Bickman, L.(1996)	NM					C3			C3	NM			NM			NM		
Nutter, M., et al (1995)	NM					C3			C3			C1			C3			C3
Webb, J., et al. (1999)			C3			C3			C3			C3			C3			C3
Goodman, R,et al. (1996)	C1					C3			C3	C1			C1					C3
Constantine, N.and Curry, K.(1998)	NM			C1			NM					C1			C1			C1
Kirby, D., et al.(1997)	NM			C1			C1					C1			C1			C1
Biglan, A.,et al.(1996)	NM				C1			C1				C1			C1			C1
Mertens, D.(1997)	C1			NM			C1					C3	NM					C3
Collins, C.,et al.(1999)	C1				C3			C3				C3			C3			C3
Herman, S.(1997)	C1			NM			C2				C1		NM					C1
Peyrot, M.,et al.(1994)	C1				C1				C1	C1					C3			C3
Dagenais. F.(1978)	C1			C1			C1			C1			C1			C1		
Patrick, D.,et al. (1999)	C1			NM			C1			C1			NM			C1		
Sussman, S.,et al. (1997)	C1			C1			C1			NM			NM			NM		
Sandler, I.,et al.(1992)	C3				C3				C3	C1			C3			C3		
<b>Note:</b> NM in a box indicates the theory origin was not mentioned in the evaluation report. C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based. The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).																		



## LEVELS OF STAKEHOLDER THEORY

In the application of the Classification Matrix to evaluation reports, stakeholder levels were at times important to delineate, particularly with regards to theory origins. For example, Brekke (1987) provides a good example of at least two levels of stakeholder theory origin. One group were those responsible for the Training in Community Living Model upon which the program was based, and a second group were those responsible for implementing the program based on the Model. This distinction in program

levels at which an evaluation can take place, either at the program level, the project level, or the element level. To illustrate, consider the program level to be equivalent to a national undertaking, the project level to be the implementation of that program at a particular site and the element level to be the investigation of a particular component of the project. Certainly it is expected that the level of stakeholder group involved in theory articulation would influence the resultant program theory.

## EXPERIMENTAL EVALUATIONS VERSUS THEORY-BASED EVALUATIONS

In the review of these evaluation reports it was evident that a proportion of recognised theory-based evaluations are of experimental programs, or programs designed and implemented to put a theory to the test (Biglan et al., 1996; Cohen and Rice, 1995; Kirby et al., 1997; Patrick et al., 1999; Pentz et al., 1990; Sandler et al., 1992; and Sussman et al., 1997). Biglan et al. (1996) term their evaluation of an experimental program an 'experimental evaluation'. The evaluators are also the program designers and were responsible for overseeing program implementation. They have conceptualised the theory driving the program and developed the overall evaluation strategy in the initial stages of program development, not as part of the evaluation process.

'Experimental evaluations' are by their nature heavily theory-driven. Yet should a distinction be made between program evaluation undertaken as part

of the data collection stage of a social research project on the one hand, and the program evaluation developed to address a problem or need in society on the other? The main difference seems to lie in the timing of the program theory articulation. The program theory of an experimental program, as conceptualised by Biglan et al. (1996), is articulated prior to program development by the social researchers. The social researchers also drive the program development, program implementation and evaluation. Alternatively, the ‘non-experimental’ theory-based evaluation involves a theory articulation process, generally undertaken post-program implementation, in which a variety of information sources may be considered. The evaluators and the program stakeholders are often different people, unlike the dual or multiple role status of ‘experimental evaluation’ evaluators.

#### MEASURE OF PROGRAM THEORY USE CLASSIFICATIONS

The final step in the classification of type, degree and contextualisation of program theory use adopted in each evaluation is the determination of numerical measures based on the Classification Matrix data for each evaluation. This section focuses on the logic underpinning the values assigned to designate program theory use by type, degree and level of contextualisation. Each of the 21 evaluations classified above were scored with regards to program theory use by type, source, degree and level of contextualisation.

#### PROGRAM THEORY USE CLASSIFICATION SCALING

As mentioned earlier in this chapter, in the selection of evaluations for this research project, an attempt was made to include studies which varied in the degree of program theory use with regard to both program implementation theory and program causative theory. In consideration of the original design proposed for this research with regards to case selection by theory type (implementation or causative theory) and degree of program theory use, it was decided that two scores would be given for program theory use for each case, one for program implementation theory use and one for causative theory use.

A score of '0' to '6' is possible for each source of program theory (Social Science Theory, Evaluator Theory, and Stakeholder Theory) within each type of theory use (Implementation Theory and Causative Theory). Simply, a score of '0' has been awarded for 'No or Negligible' theory use. A score of '1', '2' or '3' has been given for 'Minimal' program theory use depending on the level of contextual consideration included in the theory ('C1', 'C2', or 'C3'), while a score of '4', '5' or '6' has been allocated for 'Substantial' program theory use, again depending on the level of contextual consideration. The scoring system is represented in Table 3.9. The scores for all 21 evaluations classified above with regards to program theory use by type, source, degree and level of conceptualisation are given in Table 3.10.

**Table 3.9      Program Theory Use Score  
Guide for Case Study  
Selection**

<b>Program Theory Use Score</b>		
<b>Level of Program Theory Use</b>	<b>Level of Contextual Consideration</b>	<b>Score</b>
No Mention and No/Negligible	N/A	0
Minimal	C1	1
	C2	2
	C3	3
Substantial	C1	4
	C2	5
	C3	6

## CONCLUSION

The five sections of this chapter have concentrated on furthering understanding of program theory-based evaluation, and elaborating the conceptual platform upon which this study is based. Within this chapter attention has been given to the theory-based evaluation process, the development and application of a data collection matrix to guide the collection of information considered important to understanding a theory-based evaluation, the presentation of a model with text to illustrate the temporal flow of the theory-based evaluation process, and the Core Model and further elaborations of the model for testing in this study have been detailed.

The final section describes the data collection matrix. This classification exercise has provided useful insight into different approaches to theory-based evaluation. Some important sources of variability contributing to the diversity evident in theory-based evaluation practice have also been highlighted, including: the level of contribution of social science, stakeholders, and evaluators to the elaboration of program implementation theory and program causative theory, the level of theory included in the evaluation process; and the consideration of the contextual conditions related to the program. The review has collected a number of examples of theory-based evaluations and applied a systematic analysis to the program theory classification process. A trial of this method has been described and, given the results, it can now be used to measure the extent of program theory use in the evaluation studies which form the primary source of data for this research study.

The consideration of the theory-based evaluation process, program theory development and the development of the data collection matrix described in this chapter have been necessary not only from a theoretical perspective, but also to inform practice. The development of the Process Model (Figure 3.3), Program Theory Development (e.g. Figure 3.14), the Data Collection Matrix (Figure 3.16), and the Classification Matrix (Figure 3.1) has provided guidelines for theory-based evaluation practice for evaluators, and hopefully will facilitate the conduct of quality future theory-based evaluations.

**Table 3.10 Classification of Theory-Based Evaluation Reports**

	Implementation Theory				Causative Theory			
Theory Source	Social Science Theory	Stakeholder's Theory	Evaluator's Theory	Implementation Theory Use	Social Science Theory	Stakeholder's Theory	Evaluator's Theory	Causative Theory Use
Evaluation Study	Theory Source				Theory Source			
Chen, H. et al	0	6	6	12	0	1	0	1
Brekke, J.	6	6	0	12	0	0	0	0
Bickman, L.	0	4	5	9	0	0	0	0
Milligan, S.	0	6	4	10	0	0	0	0
Pentz, M. et al	1	6	6	13	0	0	0	0
Cohen, D and Rice, J.	0	4	4	8	0	0	0	0
Bryant, D. and Bickman, L.	0	6	6	12	0	0	0	0
Nutter, B. et al	0	6	6	12	4	6	6	14
Webb, J. et al	6	6	6	18	6	6	6	18
Goodman, R. et al	0	6	6	12	0	0	6	6
Constantine, N. and Curry, K.	0	0	0	0	4	4	4	12
Kirby, D. et al	0	0	0	0	4	4	4	12
Biglan, A. et al	0	1	1	2	4	4	4	12
Mertens, D.	1	0	0	1	6	0	6	12
Collins, C. et al	1	3	3	7	6	6	6	18
Herman, S.	1	0	0	1	1	0	4	5
Peyrot, M. et al	1	1	4	6	0	6	6	12
Dagenas, F.	1	0	0	1	0	0	0	0
Patrick, D. et al	1	0	0	1	0	0	0	0
Sussman S., et al	1	0	0	1	0	0	0	0
Sandler, L. et al	0	3	3	6	0	0	0	0

## **CHAPTER IV**

### **METHODS OF THE INVESTIGATION**

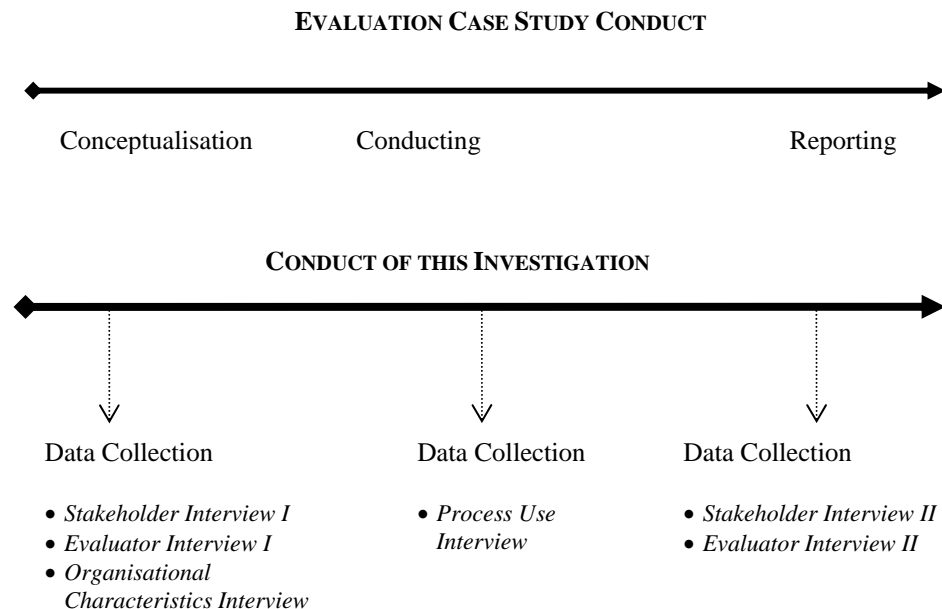
This chapter is presented in six sections. The first section is concerned with the focus and design of the study. The second describes the study in terms of the design, both the pros and cons of the chosen design and the benefits of selecting it over other study designs considered. The sources of information of the study, including the criteria for the selection of the evaluation case studies that are the focus of this investigation and described more fully in Chapter VI, are dealt with in the third section. Summarisations of the scales of the instruments developed for the study, as well as other sources of information accessed, are considered. The fourth section focuses on the procedure of the study, the strategy chosen for this investigation, justification for the adoption of this approach and the ethics guiding the study. In addition, some alternative investigation methods are considered in terms of their strengths and shortcomings for the present investigation. The conclusions are the focus of the final section of the chapter.

#### **STUDY FOCUS AND DESIGN**

The intent of this research study is to investigate the effect of the theory-based evaluation approach on the use of the information provided by such studies using this approach. Although interest in the theory-based evaluation approach has increased in the past decade, as is evident in the growth in conceptual and empirical literature addressing the topic (Lipsey et al., 1985; Reynolds, 1998; Rogers et al., 2000; Weiss, 1997b; Worthen, 1996), a review of the relevant literature indicates there is a lack of studies undertaken to investigate the influence of theory-based evaluations on evaluation information utilisation. This chapter outlines the research methods used to conduct such a study.

## DESIGN OF THE STUDY

The longitudinal case study approach was adopted for this study as it is an intensive way to look at a number of variables selected for consideration in this investigation, with a view to understanding the influence they have on the use of a program evaluation's information. The data collection was nested in each case. Data was collected from three groups; stakeholders of the evaluation studies, evaluators of the studies and representatives of the organisations responsible for the delivery of the programs evaluated. The data collection points were determined by the information required to test the conceptual model of the study presented in Chapter III. Data was collected at three points in the evaluation process of each case; at the beginning of the study, at a point midway through the study, and at a point in time distant enough from the delivery of the final report to the stakeholders to allow time for them to process the information and consider ways in which the information might be put to use. A diagram summarising the study design is presented in Figure 4.1.



**Figure 4.1: Study Design Diagram**

The case study approach enables the in-depth investigation of each program and evaluation study independently. The influences on the evaluation and the use of the evaluation information can then be compared and contrasted between the three studies, with a view to gaining insight into the influences on the use of the evaluation information. This study examines the influence of the evaluation's information from the perspective of the evaluation stakeholders. A longitudinal case study design, as adopted by Cummings (1997), was used for this investigation. A range of different designs were considered which might yield the required information. The case study design was selected as it was considered to be a better approach to identifying the use of the information of each evaluation and determining the characteristics of each case that influence that use.

The merits and disadvantages of both retrospective and a pre-post design were considered for this study. However, a decision to include a focus on process use (Patton, 1997), or the influence of the evaluation process on evaluation stakeholders, made it necessary to adopt a strategy of investigating the evaluation as it occurred. Certainly stakeholders could have been asked to consider evaluation process use retrospectively; however it is likely some detail may have been lost simply due to the inability of stakeholders to accurately recall such information months after the event. The likelihood is stakeholders would easily recall the most obvious details and tend to forget the less obvious ones. Furthermore, when considering evaluation information use retrospectively, it is difficult to attribute change to the evaluation information specifically, discriminating between the influences of the evaluation process, the evaluation information and other influential factors present in the environment (e.g. political climate) (Leviton and Hughes, 1981). Finally, staff movement both within and between public departments and organisations responsible for the delivery of the evaluations, is typical. Waiting to the end of the evaluation to interview stakeholders would, in all three cases, have meant losing the opportunity to capture experiences of many stakeholders involved in the earlier stages of the evaluations.



One primary disadvantage of the case study method is the difficulty encountered when trying to compare unique cases. In the present study, the questionnaires developed to collect data from the target populations include a number of closed-response items targeting specific information relating to the program, the evaluation, the evaluators, the stakeholders and the organisation variables. The logic underpinning the inclusion of the closed-response items is to enable analysis across cases in key areas of interest. This is a method identified by Stake and Easley (1978) as the “case survey method” (p. 46), enabling the analysis of pertinent issues across sites. Furthermore, the inclusion of interview items linked to the same scales, in interview schedules administered before the conduct of the evaluation and after the evaluation has been completed, enables a comparison of stakeholder perceptions regarding characteristics of the evaluation study and their involvement in the evaluation study before and after the evaluation study is conducted. In addition, a number of open-response items have also been included to allow subjects to respond more fully, making available further insight into their perceptions and thinking.

The independent variables in this study include characteristics of the stakeholders, the organisation, the evaluators and the evaluation in each case. In addition, the influence of the independent variables included in the study on each other, or variable interaction, is considered. The dependent variable is the use of the evaluation information.

As mentioned in the initial chapters of this dissertation, although theory-based evaluations are becoming a more commonplace practice, there is not a great deal of research detailing the influence of a program theory-based evaluation approach on the use of the evaluation information. Therefore, research in this area is more exploratory than confirmatory. As the present study is a first in this area, a case study approach has been chosen as it offers opportunity for in-depth consideration of the evaluation processes as well as the use of the evaluation information. The primary disadvantage of the case study approach is the limitation of generalisability to other studies, though the adoption of the case study survey method mentioned above addresses this disadvantage to

some degree. However, this study represents the initial research considering this area. Once particular influences are determined then research with a confirmatory approach may become more commonplace.

## **SELECTION OF THE CASE STUDIES**

The purpose of this section is to detail the processes by which the evaluation studies of three programs (Pre-Apprenticeship Program, Youth on Health Festival, Making Consistent Judgments), detailed in Chapter VI, were selected for inclusion in this investigation. In determining the study strategy it was decided the selected evaluation studies had to fit with certain pre-determined criteria. The determination of criteria for evaluation study selection enabled the comparison and contrast of evaluation use by holding particular variables constant among the three studies, thereby forcing certain evaluation study similarities. These criteria are elaborated below. Further discussion is included where the criteria could not be applied perfectly.

## **CRITERIA FOR STUDY INCLUSION**

The six criteria which guided the selection of the case studies are outlined below in priority order.

## **Extent of Program Theory Use in the Evaluations**

An important consideration upon which this investigation is based is the influence of the use of the program theory in the evaluation on the utilisation of the evaluation information. The effect of the use of a program's theory in the evaluation process could include influence of: the stakeholder and evaluator conceptualisation of the program and/or decisions regarding the program; the evaluation strategy adopted; the focus of the data collection instruments; data interpretation; and the extent to which the evaluation findings are linked to the theory underpinning and guiding the program.

Only in an ideal world would it be possible to select program theory-based evaluation studies for research focus falling within particular classifications regarding the use of program-theory by type and degree. Evaluation studies are dynamic and ever-evolving processes influenced by a myriad of variables, many of which are outside the control of the evaluator or the program stakeholders. When conducting research involving evaluation studies, some allowances must be made for that which cannot be controlled. However, identifying as many of these variables as possible at least enables the consideration of their influence in the particular case under investigation.

Early discussions were undertaken with relevant evaluation project managers and/or program stakeholders to facilitate a common understanding regarding program theory and how the theory might be put to use in the evaluation study. These discussions probably influenced evaluation project managers/program stakeholders to include program theory in their evaluation studies, but since this study was distinct from the programs themselves, the influence was not thought to be a controlling influence.

### **Duration of Evaluation Studies**

Only evaluation studies expected to be of six to eight months duration were considered for inclusion in this investigation. This time period was set to ensure the studies were of sufficient length to enable the data collection points to be independent of each other, but to reduce the problem of stakeholder turnover, which becomes more prevalent in longer evaluation studies.

### **Stakeholder Group Identification**

The evaluation stakeholders are the primary sources of information contributing to the findings of this investigation. Therefore, in the identification of appropriate studies to be included, it was important that the evaluation had an identifiable reference group or stakeholder group responsible for consulting with the evaluation team regarding the evaluation, receiving the evaluation findings, and putting the evaluation information to

use. Weiss (1983) describes a stakeholder as someone who has a stake in the program or its evaluation. For the purpose of this research, the stakeholder is considered to be concerned with the evaluation study, not the program which is the focus of the evaluation. It is the evaluation that is of particular concern in this study, not the program evaluated. There is such overlap, however, between those considered to be stakeholders of the evaluation study, and those considered to be stakeholders of the program, that it is often difficult to discern a difference.

### **Evaluation of Ongoing Program**

Only evaluation studies of ongoing programs were considered for inclusion in this investigation. Certain characteristics of an ongoing program, as opposed to a one-off program or project, may influence the use of the evaluation information (Hudson-Mabbs, 1993). For instance, formative evaluations of ongoing programs have an avenue to inform future program developments of the program, as opposed to summative evaluations of one-off programs or projects. Furthermore, stakeholders of ongoing programs with previous evaluation experience have had the opportunity to resolve issues relevant to the use of an evaluation's information that may not yet have been encountered by stakeholders of a one-off project or program. Also, stakeholders of an ongoing program, who have previous experience of evaluations tend to be more critical of evaluation practice than of evaluation information. Finally, stakeholders of an ongoing program often have some insight as to how they may put the evaluation information to use. Alternatively, stakeholders of one-off programs may not have the opportunity to use the evaluation information.

### **Program Stage of Development**

A criterion for study inclusion focused on the stage of program development. Evaluations were sought for programs which had moved beyond initial stages of implementation. Programs for inclusion were settled in their second or third implementation, or year of implementation. Most programs are run in line with the calendar year or fiscal year. Therefore, annually, there tends to be an

opportunity for reflecting on the year's practice, either formally or informally, before moving into the next year. Programs sufficiently long-standing to have developed this cycle are likely to have established a routine regarding the use of information in the development of the program and the program stakeholder's will have had an opportunity to experience this routine. The logic underpinning the decision to include this criterion relates to minimising the barriers to use of the evaluation information.

The evaluations sought for this study fell within the 'Interactive' and 'Monitoring' categories of evaluation forms presented by John Owen (2006) in his typology focusing on evaluation study forms and approaches. Although the issues typically addressed by the two forms are not mutually exclusive, the orientations of the two forms differ. The interactive evaluation form is orientated towards program improvement, while the monitoring evaluation form is orientated towards checking, refining and accountability.

### **Location of Evaluative Consultants to the Program**

The final criterion is the stipulation that the evaluative consultants must be external to the program. Evaluator location determines, to a great extent, the relationship between the program stakeholders and the evaluation team (Scriven, 1991). Internal evaluators tend to have a history with a project, the relevant stakeholders and practitioners. Experiences of the past influence current relationships and lines of communication. Furthermore, the stakeholder role of internal evaluators is two fold. Their position in the organisation offers them the opportunity to be stakeholders from the organisation's point of view. Internal evaluators are part of the organisational process, and as such represent possible hosts of the dominant prescriptive theories of the organisation. Alternatively, their position as program evaluator calls for them to be stakeholders of the evaluation. External evaluators have no such dichotomy of status. Their role is to evaluate the project as contracted and they are less likely to be influenced by either personal or "job-benefit" considerations (Scriven, 1991, p. 160).

## Other Considerations

Only evaluation studies with stakeholder groups willing to be involved in this investigation could be included. Before their permission was sought, the stakeholders of each case were informed of the commitment they would be making in terms of interview time and other requirements of this investigation. However, the population in Perth, Western Australia, from which the evaluations selected for inclusion in the study were drawn, was very limited. The criteria for inclusion in the study, and in particular, the use of program theory, substantially narrowed the number of cases appropriate for the study from which to select. None-the-less, other potential theory-based studies were likely available elsewhere.

### PROCEDURE FOR THE SELECTION OF CASE STUDIES

The process of identifying cases for this investigation took place over a six month period. A non-probability purposive sampling method (de Vaus, 1990), otherwise known as *purposeful sampling* (Patton, 1990), was adopted and cases were selected in accordance with criteria and characteristics detailed above. The purpose of this method of sampling is, "...to select information rich cases whose study will illuminate the questions under study" (Patton, 1990, p. 169). In particular, a specific type of purposeful sampling was undertaken known as *maximum variation sampling* (Patton, 2002) in which the sampling aims to capture and detail the priority theme that transects across participants or, in this case, evaluation variation. Maximum variation sampling brings to the study findings two strengths. The first is the detailed description of each case, useful for documenting the unique aspects of each case. The use of program theory in the evaluation process is the sampling aspect of interest in this study. The second is the latitude to consider shared patterns that emerge regardless of case variation in terms of the sampling aspect (Patton, 2002).

Evaluation studies included in this investigation were determined prior to their conduct. Evaluations were identified as possible choices for inclusion in this investigation through informal and formal professional networks. In each case,

consent was given by the relevant evaluators and program authorities before the investigation of the evaluation began. A number of studies were considered, though for a range of reasons found not to be appropriate cases for inclusion in this investigation.

Meetings were held with either the primary evaluator or a person responsible for managing the evaluation project, prior to selection of the evaluation study. In these meetings, the expected use of program theory in the evaluation conceptualisation and conduct was discussed. In particular, what was meant by program theory, program implementation theory and program causative theory was given to ensure a common understanding of these concepts.

It is often difficult to determine particular circumstances influencing the conduct of an evaluation study prior to its undertaking. This was found to be the case in attempting to pre-determine absolutely the type and degree of program theory use in an evaluation study. Therefore it was decided the evaluation studies included in this study would be considered in accordance with the characteristics of evaluation program theory use identified in the Classification Matrix developed for this work. The purpose of the Classification Matrix was to enable the identification of variability in the program theory use by type, degree, source and contextualisation. The inclusion of studies with identified variation in program theory use facilitates the comparison and contrast of the influence of program theory use in the evaluation study process on the use of the evaluation information.

Once an evaluation was determined to be a likely candidate for inclusion in this study, the primary stakeholder or stakeholder group were contacted to arrange a meeting with a view to discussing: the evaluation; the intent of the study; the fit of the evaluation with the criteria for inclusion of this study; what resources would be required if the evaluation was included as a case (e.g. time for interviews and other resources such as the availability of relevant documents for review); and, if appropriate, the stakeholder consent to include the evaluation in this study. A timeline linking hypothetically to the progress of an evaluation study was given to program stakeholders at this time,

identifying points at which interviews would be conducted and the groups to be interviewed. Stakeholders were then able to refer to this information when making the decision about whether to allow their program and its evaluation to be included in this research. The common reasons stakeholders of programs gave when declining the request to include their evaluation were the lack of time available to stakeholders for interviews and concerns regarding client confidentiality.

## **SOURCES OF INFORMATION**

The primary sources of information for this study were the evaluation stakeholders of the cases and those responsible for undertaking the evaluations. Structured interviews were conducted with member of both of these groups on multiple occasions. Due to the complex nature of program evaluation utilisation however, additional valuable insight into the program, organisation, evaluator, evaluation report and audience characteristics was gained through other sources of information including document reviews, observations and informal communications. The selection of the sample groups for the structured interviews and the other sources of information, are described below.

### **INTERVIEW GROUP SELECTION FOR STRUCTURED INTERVIEWS**

Purposive sampling (de Vaus, 1990) is a non-probability method of sampling and was used for the selection of the interview groups. The intent of the interview schedules is to gain insight into the evaluations, the programs they focus on, and those involved and/or with potential to be influenced by the evaluations. Table 4.1 summarises the interview groups by the number of individuals interviewed by program and sample group. For example, due to staff turnover the interview groups for the Stakeholder I Interview and the Stakeholder II Interview varied slightly. In most cases stakeholders who had left their positions, were not interviewed, and those who had replaced them in their position were interviewed instead. Table 4.2 summarises the final sample group sizes by program and interview schedule.



**Table 4.1: Number of Unique People Interviewed by Program**

PROGRAM	SAMPLE GROUP				
	Organisation Representatives	Evaluators	Program Stakeholders	Total	
	Pre-Apprenticeship Program	5	2	15	22
	Making Consistent Judgments Program	6	6	15	27
	YohFest Program	5	2	16	23
	Total	16	10	46	72

**Table 4.2: Number of Interviews Conducted**

PROGRAM		Organisational Representatives	Evaluators	Program Stakeholders	Total-
	Pre-Apprenticeship Program	5	2	13	20
	Making Consistent Judgments Program	6	6	12	24
	YohFest Program	5	2	10	17
	Total	16	10	35	61

## ORGANISATIONAL REPRESENTATIVES

Interviewees for the Organisational Environment Interview Schedule were representatives of the organisation not involved with the evaluation included as a case in this study, or with the program evaluated. They were selected as representatives knowledgeable of the organisation environment and general practices of the organisation. The purpose of this interview schedule was to gain insight into the organisation.

In the case of each of the programs, the responsible organisation was identified and individuals from a range of organisation levels identified as potential interviewees by a key member of the evaluation stakeholder reference group. In consideration of those identified as organisational interviewees, it was recognised that there was a need to include representatives from a variety of hierarchical levels within the organisation with a view to gaining a balanced insight into the environment and practices of the organisation.

Potential interviewees were then contacted by telephone and informed of the details of the research study, how they had been identified as a possible interview candidate, the logic and purpose of the interview schedule, and the amount of time required for the interview. All interviewees agreed to be interviewed. Six representatives were interviewed for the Pre-Apprenticeship program, five for the Making Consistent Judgments program, and five for the Youth on Health Festival (YOHFest) program.

Although two of these programs are delivered by the Department of Education and Training, the Preapprenticeship program is delivered through the training division and the Making Consistent Judgments delivered through the education division. Although considerable work has been undertaken to amalgamate the two, at the time of this study the two divisions operated as predominately independent bodies, each with its own management structure. Organisational representatives for training had little or no history of involvement with education. Similarly, organisation representatives of the

education division had little or no history of involvement with training. The insights of the two groups into the organisation were by for all intents and purposes of two different organisations, albeit pinned at the top.

Five representatives of the YOHFest program, also the five members of the management board for the organisation, were the only candidates appropriate for the Organisation Environment Interview. As the organisation is small, the available number of those appropriate to interview was limited. As with the other two programs, following a meeting in which the details of the interview schedule were made clear, a primary stakeholder of the program provided contact details of potential interview candidates. Candidates were then contacted with a view to gaining their permission to be interviewed. All five members of the board agreed to be interviewed.

#### SELECTION OF THE STAKEHOLDER GROUPS

A reference group was established for each evaluation study, as a criteria for inclusion in this research study. In all three studies, members of the primary stakeholder/reference group were contacted to be interviewed. All consented to be interviewed. These reference groups provided the core for the sample of the stakeholders to be interviewed. The inclusion of primary stakeholders was particularly important as in all cases they were the recognised gatekeepers of the programs. In addition to interviewing primary stakeholder reference group members, other stakeholders pertinent to the evaluation, identified in meetings with primary program stakeholders, were contacted with a view to gaining their consent to be interviewed. Again, all contacted stakeholders consented to be interviewed. They were interviewed three times: at the beginning of the evaluation (Stakeholder Interview Schedule I), midway through the conduct of the evaluation (Process Use Interview Schedule), and following the release of the final evaluation report (Stakeholder Interview Schedule II). However, the members of the interview groups did not remain static between interviews. Due to staff turnover. The interview groups changed slightly from the Stakeholder I Interview to the Process Use Interview, and again for the Stakeholder II Interview.

At the time of the first interview, stakeholders were told of the need to conduct further interviews with them at a later stage. The intent, focus and timeframe of the Process Use Interview Schedule and the Stakeholder Interview Schedule II was discussed and verbal permission sought to contact them, at a later date, with a view to undertaking the two interviews. All stakeholders gave their consent to participate in further interviews. In all cases, the Stakeholder I and II Interview Schedules were conducted in person while the Process Use Interviews were very short in duration (no longer than ten minutes) and were conducted by phone.

#### SELECTION OF EVALUATOR GROUPS

Those responsible for undertaking the evaluation study are a rich source of information regarding characteristics of the evaluation study process. However, they bring to the study their own influence, which in turn has influence on how the study is conducted, perceived and its impact. Two interviews were conducted with the evaluators of each case. The first, the Evaluator Interview Schedule I, was conducted at the beginning of the evaluation process, while the second, the Evaluator Interview Schedule II, was conducted following the release of the final evaluation report. In the selection of the evaluator interview group for each case, the aim was to include all of those involved in the conduct of the evaluation, not just the primary evaluator or evaluation manager. In all three cases, the evaluation manager was identified early in the study through professional channels. However, permission to interview the evaluators was first sought from the primary stakeholders or reference groups of each program. The evaluation manager was then contacted to gain their consent to be interviewed for the study. At this time, further members of the evaluation team were identified and permission sought to contact them, with a view to including them in the evaluator interviews. It was important to gain the permission of the primary evaluators, as in all three cases the evaluations had been contracted-out to agencies specialising in evaluation and research. The primary stakeholders or reference group of the evaluations were not in a position to determine the actions of the evaluators.

The primary evaluators of the YOHFest and Pre-Apprenticeship programs both identified one other person who had the responsibility of working to assist them in undertaking the evaluations. With regards to these two programs, both were interviewed for both of the Evaluator Interview Schedules.

The Making Consistent Judgments (MCJ) program evaluator group selection varied slightly from that of the other two programs. In discussions with the primary stakeholders of the group, they identified themselves as having undertaken a considerable amount of the evaluation conceptualisation and instrument development, as the conditions program funding required specific information from the evaluation. Furthermore, the program represented a major change in the manner by which such programs had been delivered previously by the organisation; therefore they had a particular interest in the focus of the evaluation. They perceived the evaluation to be a step in the development of the program. So, in addition to the primary evaluator and the two evaluators he identified as having responsibilities regarding the MCJ program evaluation, three evaluation stakeholders with heavy involvement in the evaluation process were included in the evaluator interview group. A total of six were included in the MCJ program group interviewed for the Evaluator Interview group.

The members of the evaluator groups interviewed for the Evaluator Interview Schedule II varied slightly to those interviewed for the Evaluator Interview Schedule I. With regards to the evaluators of the YOHFest program evaluation, the duties of the original primary evaluator had been assumed by another experienced evaluator. In this case, the members of the original evaluator group were interviewed, as well as the new evaluation member.

The number interviewed for the Evaluator Interview Schedule II group for the MCJ program evaluation was slightly less than for the Evaluator Interview Schedule. In discussions with program stakeholders originally included in the evaluator interview group, it was decided they were not good subjects for the second evaluator interview schedule, as the heavy involvement they had in the

early stages of the evaluation had not continued. Therefore, they declined to be interviewed a second time. Although two primary evaluators had remained involved, the third evaluator, who had been assisting with the data analysis for the study, had moved on. Therefore, only two members were included in the interview group for the MCJ program Evaluator Interview Schedule II interviews.

## **DOCUMENT REVIEW**

A review of documents relevant to the program of each case and each evaluation was undertaken with a view to gaining insight into the program and the evaluation. Program and policy documents, in each case, enabled insight into the history and logic necessary for detailing the descriptions of each program included in this study (Chapter VI). With regards to the Making Consistent Judgments and Pre-Apprenticeship programs, reports of previous evaluations were available.

Documents relevant to each evaluation included reports of previous evaluation studies, original evaluation proposals for the current evaluations, notes of meetings for the purpose of evaluation conceptualisation and program theory elaboration, interim evaluation reports, drafts of final evaluation reports, and the final evaluation reports. This information helped to provide further insight into the process of each evaluation and illuminated influences with regards to each evaluation and its use.

## **OBSERVATION**

Opportunities for observation were only available with regards to the YOHFest program evaluation. One meeting of primary stakeholders and evaluators for the purpose of program theory elaboration and evaluation conceptualisation, was held. This was an opportunity to gain insight into the stakeholders' thinking regarding the program and their expectations in terms of the information they wanted from the evaluation. A second opportunity for observation was taken in attendance of the final of the festival. YOHFest is a

program where high school groups compete through the medium of drama and art pieces developed around a central theme considered supportive of positive lifestyles. In addition to being entertaining, attendance of the festival was an opportunity to capture in a discreet observation, a snapshot of the program substance.

## **INFORMAL COMMUNICATIONS**

Communications with the evaluators and stakeholders of each case was ongoing, and not strictly limited to agreed upon interview sessions. The inclusion of informal communications as a study data collection procedure allowed the researcher the latitude to follow through with investigation of characteristics identified as impacting upon evaluation utilisation in the progress of the interviews, document reviews and observations. Furthermore, in instances where stakeholders originally interviewed at the beginning of the study had moved into positions where they were no longer involved with the program, further structured interviews were at times not appropriate. However, as previous stakeholders of the program and evaluation they often remained in touch with the program informally. Informal conversations with them often provided useful information regarding the evaluation or confirmed information gathered through other sources. Finally, with regards to the YOHFest program, one stakeholder who had previously been heavily involved with the program was identified. Although this person was no longer involved with the program they were very knowledgeable about the program history and the influence of one significant sponsor body. A lengthy conversation took place with this person for the purpose of gaining further program insight.

## **DATA COLLECTION INSTRUMENTS**

The data collection methods used for this study include a review of pertinent documents, structured interviews, observation and informal communications. Each of these methods is detailed more fully below.

## **STRUCTURED INTERVIEW SCHEDULES**

Structured interviews were the main data collection instruments for the study. The six interview schedules developed for the study are detailed below. They include the Organisational Characteristics Interview Schedule (Appendix A), the Stakeholder Interview Schedules I and II (Appendix B), the Evaluator Interview Schedules I and II (Appendix C), and the Process Use Interview Schedule (Appendix B). Each is detailed below. Table 4.3 was originally developed to enable stakeholders of evaluation studies considered for inclusion in this investigation, some insight regarding what consenting to allow their evaluation and program to be studied would entail. It is a generic representation of the timing of the interviews in this study, which summarises details of each interview schedule including focus, target subject group and timing.

All six interview schedules contained a mix of open response and closed response items. The information gathered from the open response items was expected to add insight to the quantitative data. A few items requiring a dichotomous response have been included, generally as screening items to determine the appropriateness of asking a particular set of questions of an interviewee. With the exception of the four scaled response items included in the Evaluator II Interview Schedule, all other scale items are anchored to a continuum from 1, representing “Not at All” to 5 representing “A Great Deal”. The scales of the items included in the Evaluator Interview Schedule II will be further detailed in the description of the instrument later in this chapter. A total of 24 scales have been developed for the study. The scales and sub-scales of each interview schedule are summarised below by interview schedule. A fuller description of the instruments is undertaken in Chapter V together with the reliability analysis of the scales.



**Table 4.3: Descriptions of Study Interview Schedules**

<b>Organisational Characteristics Interview Schedule (approximately 30 minutes)</b>	
* FOCUS:	The characteristics of the organisation responsible for program delivery.
* INTERVIEWEES:	5-6 people from different levels and areas of the organisation.
* WHEN:	Interviews conducted just prior to the evaluation.
<b>Evaluations Team Interview Schedule 1 (approximately 30 minutes)</b>	
* FOCUS:	The characteristics of those responsible for undertaking the evaluation study.
* INTERVIEWEES:	The evaluation team.
* WHEN:	Interviews conducted just prior to the study commencement or in the early weeks of the study.
<b>Stakeholder Interview Schedule 1 (approximately 30 minutes)</b>	
* FOCUS:	To gain insight into the training, experience and evaluation study expectations of those linked to the evaluation.
* INTERVIEWEES:	Those primarily involved with the evaluation and/or influential to the evaluation process.
* WHEN:	Interviews conducted in the early weeks of the evaluation study.
<b>Process Use Interview Schedule (approximately 15 minutes)</b>	
* FOCUS:	The ways in which the evaluation study process may have influenced those involved.
* INTERVIEWEES:	Those primarily involved with the evaluation and/or influential to the evaluation process.
* WHEN:	Interviews administered by telephone twice during the evaluation study process.
<b>Evaluator Interview Schedule 2 (approximately 30 minutes)</b>	
* FOCUS:	The process of the evaluation study and use of the evaluation information.
* INTERVIEWEES:	The evaluation team.
* WHEN:	Interviews conducted at the end of the study.
<b>Stakeholder Interview Schedule 2 (approximately 30 minutes)</b>	
* FOCUS:	Perceptions of the evaluation study process and use of the evaluation information.
* INTERVIEWEES:	Those primarily involved with the evaluation and/or influential to the evaluation process.
* WHEN:	Interviews conducted following the release of the final report.

## ORGANISATIONAL CHARACTERISTICS INTERVIEW SCHEDULE

The instrument included in this study to assess organisational characteristics is a modified version of the Readiness for Organisational Learning and Evaluation Instrument (ROLE) developed by Preskill and Torres (2000b) to determine the extent to which this supportive infrastructure is present in an organisation. It is a diagnostic instrument developed to determine the presence of organisational aspects supportive of evaluative inquiry and organisational learning. These aspects, or characteristics, are organisational culture, forms of communication, and systems and structures present to support evaluative inquiry.

In the present study the particular organisational factors chosen for inclusion are considered to be indicators of the ability of the organisation to learn. The scales developed focusing on the organisation representatives' perceptions are briefly described below. The items of the interview schedule contribute to the five scales developed for the interview.

*Characteristics of the Environment of the Organisation scale.* The purpose of the scale focusing on the environment of the organisation is to gain staff views of the organisation's environment or culture. The scale items focus specifically on information regarding the extent to which the organisational environment is supportive of staff learning.

*Characteristics of the Leadership Practice of the Organisation scale.* The items of this scale seek information from the respondent regarding the extent to which managerial staff model behaviours and practices characteristic of staff employed by a learning organisation. If managerial staff model these behaviours, it is more likely that those they supervise will also adopt them. Leaders of an organisation are key in building and sustaining environments supportive of learning (Preskill and Torres 1999b).

*Characteristics of the Personal Practice scale.* The purpose of this scale is to gather information from the interview group to determine the extent to which

behaviours characteristic of those employed by a learning organisation are occurring. For an organisation to be a 'learning' entity it is important that the environment within which the staff work is one that encourages and facilitates learning new knowledge and practical application of that which is learned. However, if on an individual basis practices characteristic of a learning organisation are not occurring, then the ability of the organisation to learn is interrupted.

*Characteristics of the Organisation's Use of Information scale.* The items of this scale focus on the perceptions of the interview group regarding the information collection, dissemination and use practices of the organisation.

*Characteristics of the Organisation's Teamwork scale.* The fifth scale of the interview schedule is concerned with the perception of the interview group regarding the characteristics of teamwork present in the organisation.

#### STAKEHOLDER INTERVIEW SCHEDULES

The focus of this investigation is on the use or influence of the evaluation and its information, and on influential factors. As the primary users of the evaluation information, stakeholders are a rich source of information regarding the characteristics of the program, evaluation, evaluators, the organisation, and themselves. Stakeholders responsible for the delivery of the program are considered to be in a better position to provide insightful information pertinent to the evaluation than the clients or recipients of the program, or program stakeholders. For the purposes of this study, a stakeholder is considered to be someone with an investment in, or influence on, the evaluation study, as the evaluation study is the focus of this investigation rather than the program evaluated.

The scales of the two stakeholder interview schedules are briefly summarised below. Six scales have been included in the items of Stakeholder Interview Schedule I, while nine scales have been included in the items of Stakeholder Interview Schedule 2.

## Stakeholder Interview Schedule I

The purpose of the Stakeholder Interview Schedule I was to gather information from primary stakeholders regarding their: commitment to the program; commitment to the evaluation study; expectation regarding their involvement in the evaluation; training and experience with social research and evaluation; perception of the organisation; and experiences with the use of information in relation to the case study program. Ten stakeholders of the Pre-Apprenticeship program, 13 stakeholders of the YOHFest program and 13 stakeholders of the Making Consistent Judgments program were interviewed for the Stakeholder Interview Schedule I.

*Commitment to the Program scale.* This scale includes 13 items to determine stakeholder commitment to the program. Many of the items have been adopted from an instrument developed by Cummings (1997) to determine stakeholder commitment to the program. The facets of stakeholder commitment dealt with by the scale items include the type and level of stakeholder involvement with the program, the stakeholder's perception of the need for the program, the stakeholder's perception of program philosophy and efficacy of program delivery, and the level of stakeholder support or opposition to the program.

*Stakeholder Commitment to the Evaluation Study scale.* Five items have been included in the interview schedule to form the *Stakeholder Commitment to the Evaluation Study* scale. The items of the scale focus on stakeholder's perception of the utility of the evaluation study, or the extent to which they expect the evaluation is suitable or adaptable to an end, stakeholder's perception of the need for the evaluation study, and stakeholders' support of the evaluation study.

*Stakeholder Expectation of Involvement in the Evaluation Study scale.* The items included focusing on stakeholder's anticipation regarding involvement in all stages of the evaluation study, have been adapted from the work of Cumming (1997). Cummings included a version of these items in his *Post-Study Questions*. Here they have been adapted to gather information pre-study

to gain insight into stakeholders' expectations regarding involvement in the evaluation study. Similar items, with changes to tense of the items, have also been included to gather similar information from the stakeholders of the evaluations included in this study, following the dissemination of the relevant evaluation's final report (Stakeholder Interview Schedule II), with a view to comparing and contrasting the expectations of stakeholders regarding involvement in the evaluation study and their perception of the extent to which they were actually involved. It is expected that the more stakeholders expect to be involved in the evaluation, the more likely they will be involved.

Items included in the scale focus on stakeholder involvement in; planning, conduct, communication, elaboration of the program logic, dissemination of the findings, and development of the recommendations. In addition, two items have been included to assess the extent to which the stakeholder's involvement in the evaluation study will be expected to gain insight into the influence the expectations of the organisation have about stakeholders' anticipation regarding their involvement in the evaluation.

*Stakeholder Characteristics- Education and Training in Social Research and Evaluation scale.* The opportunity the stakeholder has to attend to research and evaluation information, to reflect upon the information and apply the information, represent the Opportunity of Stakeholders to Use Information variable in this study. Items focus on the extent to which respondents are able to read social research reports, consider the implications for their work, and apply the information to their work. These items were developed to gather information regarding stakeholders' avenue for social research information use. Four scaled response items together form the scale. It is expected previous education and experience with social research and evaluation will have a positive influence on stakeholder openness to using social research and evaluation information. The scale items focus on stakeholders' perception of their educational background and training or experience with regards to social research and evaluation, and their perception of evaluation.

*Opportunity of Stakeholders to Use Information scale.* The six items of this scale are concerned with the stakeholders' perception of the organisational culture within which they work. The logic underpinning the inclusion of the items is that for use of information to occur both conceptual and operational opportunities to attend to, and apply, the information must be available to the individuals. The items ask the extent to which respondents are "able" to read reports, consider the implications of study information and apply study information to their work, again recognising the powerful influence of the organisation culture and norms on the access and use of research and evaluation information.

*Stakeholder Commitment to the Organisation scale.* The logic linking stakeholder commitment to the organisation with use of the evaluation information is an extension of the work of Cummings (1993). The scale is composed of two sub-scales: *Perception of Organisation Direction and Need* sub-scale; and *Perception of Organisation culture, Capacity to Change, and Personal Efficacy* sub-scale. Adopted from the work of Preskill and Torres (1999b), the six items to determine stakeholder commitment to the organisation, are concerned with stakeholders' perception of: the strategic direction of the organisation; the need for organisation services; the efficacy of organisation service delivery; the culture of the organisation; the organisation's capacity to change; and their personal efficacy in relation to the organisation.

## **Stakeholder Interview Schedule II**

The Stakeholder Interview Schedule II was developed to gather information from the stakeholders of the evaluations that have been included in this longitudinal study, focusing on the characteristics of the evaluation studies and their influence on the use of the evaluation information. The interview schedule was administered post-evaluation study, following the release of the final reports to the stakeholders. The items of the interview schedule focus on stakeholders' perspectives of: the evaluation study characteristics; the influence of the evaluation information; the characteristics of the evaluators

responsible for undertaking the evaluation studies; the characteristics of the environment influencing the evaluation process; their personal involvement in the evaluation studies; the use of the program's theory, logic or rationale in the evaluation; the influence of their involvement in any work to understand or elaborate the program theory, logic or rationale; and their use of the final evaluation reports and the evaluation information.

Stakeholders interviewed for the Stakeholder Interview Schedule I were interviewed for the Stakeholder Interview Schedule II, as available. In some cases stakeholders interviewed for the first interview schedule had moved on to other positions and were no longer involved with the program. Details regarding changes to the Stakeholder Interview Group have been detailed in the Interview Group Selection for Structured Interviews section of this chapter. Nine scales are contained within the items of the interview schedule. These scales are summarised below in the pertinent sections.

*Evaluation Study Characteristics scale.* This scale is comprised of two sub-scales: *Characteristics of Commitment to Evaluation Study- Post sub-scale*, and the *Characteristics of the Evaluation Study sub-scale*. The first focuses on stakeholder perceptions regarding characteristics of the evaluation study. The second is intent on gathering information from stakeholders regarding their perceptions of the influence of the evaluation study. The items of each sub-section are summarised below.

The *Commitment to the Evaluation Study sub-scale* focuses on stakeholder perceptions of the evaluation need, usefulness of the evaluation information, usefulness of the evaluation information to the organisation responsible for program delivery, appropriate for the stage of the program's life, and the extent to which the stakeholder supports the evaluation study. The items of this sub-scale have also been included in the Stakeholder I Interview Schedule. The inclusion of these items both pre- and post- study, with minor changes to item wording in terms of tense, allows for comparison and contrast of stakeholder expectations and perspectives prior to the conduct of the

evaluation study, with their perspectives of the study following dissemination of the final report.

The *Stakeholder Perceived Characteristics of the Evaluation Study sub-scale* focuses on stakeholder perceptions of characteristics of the evaluation study including: the quality of the data collected in the study; the study timeliness for any pending decisions; the extent to which the study findings are politically acceptable; the agreement of the study findings with other sources of information; the influence of their role in their use of the study information; the appropriateness of the evaluation study methods; the extent to which the study findings are supportive of the program; the extent to which the study findings are critical of the program; the extent to which they agree with the study findings; and the extent to which they believe the study findings are relevant.

*Influence of the Evaluation Findings scale.* Five scaled response items focusing on the influence of the findings on stakeholder's thinking regarding: the program; issues beyond the program; any program decisions; decisions beyond the immediate program; and support for views or positions about the program which people already held from the *Influence of the Evaluation Study Findings scale*.

*Evaluation Team Characteristics scale.* It has been suggested by many authors that characteristics of the evaluator have influence on the use of the information from an evaluation study. This scale of the interview schedule contains seven items to gain insight into stakeholders' thinking regarding the evaluators' openness to listening to the ideas of stakeholders regarding the evaluation, and capacity to incorporate these ideas into the evaluation process. The items focus on the extent to which stakeholders felt evaluators: were approachable; were protective of their ideas for the evaluation; were tolerant of change; were open to stakeholder's ideas regarding the evaluation; were competent to undertake the evaluation study; competently conducted the evaluation study; and endeavoured to develop relationships with stakeholders.



*Characteristics of the Evaluation Environment scale.* Three scaled response items were developed to gather information from stakeholders regarding their perception of the evaluation environment. The logic underpinning the items is focused on understanding any influences constraining the evaluation process. The items focus on stakeholders' perceptions regarding the extent to which the evaluation was able to proceed as planned, the extent to which evaluators were given sufficient latitude in undertaking the evaluation, and the extent to which evaluators were free to undertake the evaluation.

*Involvement in the Evaluation Study (Post) scale.* For the purposes of this study relevant data has been collected both pre- and post- evaluation study conduct, with a view to gaining insight into discrepancies between stakeholder expectations regarding their involvement in the evaluation study and the extent to which they were actually involved. It is expected that the influence of any discrepancy between the two and their level of actual study involvement, will influence their use of the evaluation information. The nine items which comprise the *Evaluation Study Involvement (Post) scale*, have also been included in the *Stakeholder Expectation of Involvement in the Evaluation Study scale* and included in the Stakeholder I Interview Schedule, with minor changes to item wording regarding tense.

*Influence of the Use of Program Theory in the Final Evaluation Report scale.* The six items of this scale were developed to gain insight into stakeholders' perception of use of program theory in the evaluation report and the influence of the program theory in the evaluation on their thinking regarding the program, practice and any program decisions they have made or expect to make. The logic for the inclusion of these items is that an understanding of the stakeholder's perception of the program theory use in the evaluation is an important precursor to interpreting the influence of the program theory.

*Stakeholders' Perspective of the Influence of their Involvement in the Program Theory Elaboration Process scale.* The three items of this section were developed to gain insight into stakeholders' perspective of the influence of their involvement in the program theory elaboration process. The items deal

with how the involvement influenced their thinking about the program, the practice, and any decisions they have made, or expect to make, regarding the program.

*Use of Evaluation Study Information scale.* The 26 items which form the *Use of Evaluation Study Information scale* are divided into two sub-scales, each with 13 items, the *Likelihood of Use sub-scale* and the *Importance of Use sub-scale*. The items of the sub-scales focus specifically on stakeholders' perceptions of conceptual, instrumental and strategic influence of the evaluation information, regarding the program that is the subject of the evaluation and other avenues of impact. They deal with 13 manners in which the stakeholders may have put the evaluation information to use, both formally and informally. The *Likelihood of Use* sub-scale focuses on how likely the particular manner of use is to occur and the *Importance of Use* sub-scale is concerned with how important each manner of use is. The items of each sub-scale are the same with only slight changes to the wording where *likelihood of use* is substituted for *importance of use*.

#### EVALUATOR INTERVIEW SCHEDULES I AND II

Evaluators are, for the purposes of this investigation, considered to be anyone responsible for managing or undertaking either all tasks or a specific task of the evaluation process. Although the primary evaluator or evaluation manager has considerable opportunity to influence the evaluation, other members of the evaluation team also have an avenue through the undertaking of their duties. The decision to interview all layers of the evaluation team was made to account for the potential of the evaluation team to influence the evaluation individually, as well as collectively. Two interview schedules were developed to gather information from the evaluators of each case, the Evaluator Interview Schedule I and the Evaluator Interview Schedule II. The scales of both are summarised in the sections below.

## Evaluator Interview Schedule I

Two scales have been included in the interview schedule items: *Evaluator Perception of Evaluators' Practice scale* and the *Evaluator Perception of Stakeholder Involvement scale*. Both scales and the items contributing to them are summarised in the pertinent sections below.

*Evaluator Perception of Evaluator Practice scale.* 13 items together form the *Evaluator Perception of Evaluators' Practice scale* of the Evaluator Interview Schedule I. It is expected this insight into the preferred practices of the evaluator will give some insight into their own practice, in turn influencing the evaluation process. Items of this scale focus on: the extent to which evaluators feel the program stakeholders' ideas and opinions should be taken into account by the program evaluator in the evaluation process; the extent to which evaluators feel the context of the program should be taken into account by the program evaluator in undertaking the evaluation; and the extent to which evaluators consider the program theory, logic or rationale should be taken into account by the program evaluator in the undertaking of the evaluation.

*Evaluator Perception of Stakeholder Involvement scale.* Item 22 a-g forms the scale concerned with Evaluator Perception of Stakeholder Involvement in the Evaluation Process Scale. Adopted from the post-study interview schedule of Cummings (1997) focusing on the use of evaluation information as perceived by program stakeholders, the item contributes to a scale included to gain insight into evaluators' perceptions of stakeholder involvement in a range of activities linked to the evaluation process. It is expected that evaluators with a predilection towards stakeholder involvement in the evaluation process will be most likely to conduct evaluations in accordance with their thinking. The items focus on stakeholder involvement in the evaluation conceptualisation in determining the evaluation focus, program theory or logic elaboration, evaluation objectives, target information, information sources, manner of data collection, and reporting of the evaluation information.

Two further instruments developed for this study were also included in the Evaluator I Interview Schedule. The first is the Evaluator Perception of Evaluator Practice: Instrument I, the second is the Evaluator Perception of Evaluator Practice: Instrument 2. Both of these instrument are detailed more fully in Chapter V.

### **Evaluator Interview Schedule II**

The purpose of this interview schedule is to gain insight regarding the evaluation from the perspective of those responsible for undertaking the evaluation, after it had been conducted. The Evaluator II Interview Schedule contains only one scale.

*Evaluator Perception of the Evaluation Characteristics scale.* Four items of the interview schedule contribute to the scale which focuses on gaining insight into evaluator perception regarding the evaluation process in terms of the extent to which they felt; their skills were appropriate, the evaluation team was competent to undertake the evaluation study, the evaluation team was given sufficient latitude to exercise sound professional judgment in undertaking the evaluation, and the extent to which they felt the evaluation to have been a negative or positive experience.

### **Process Use Interview Schedule**

The Process Use Interview Schedule is concerned with changes in the thinking and/or behaviour of those involved in the evaluation resulting from the learning that takes place in the evaluation process, and changes to program or organisational culture and processes. The schedule contains only one scale, which is summarised in the section below.

*Process Use scale.* Four items of the interview schedule together form the Process Use scale which is concerned with stakeholders' perception of the influence of any involvement in the evaluation on; their thinking, their

thinking about issues other than the program, any decision they were involved in making, and their practice.

## PROCEDURE

The study commenced with a review of literature relevant to the undertaking of evaluation studies, with a particular focus on the theory-based evaluation approach, the influence and use and of evaluation information, and organisational learning. The literature review was undertaken to identify pertinent gaps in empirical research to be addressed by this study, and to develop and underpin the logic upon which this study is based. Figure 4.2 is a representation of the procedure of the study.

First, pertinent literature was reviewed with the focus of the study as a guide. The process of determining the study design in respect to what information was to be collected from whom, when and how, was followed by the development of the six study instruments. Following the development of the instruments, the process of identifying appropriate cases for the study began and discussions were undertaken with relevant stakeholder groups with a view to including appropriate evaluations. Due to the difficulty of identifying appropriate cases for inclusion in this study, it was not possible to trial all instruments on one evaluation prior to undertaking actual data collection for the study. The availability of studies appropriate for this study was extremely limited. The instruments were trailed individually with stakeholder and evaluator groups appropriate for each. The responses and feedback of the trial subjects was used to refine the wording of interview schedule items to clarify meaning. Therefore, the analysis of the instruments undertaken in Chapter V is based on actual rather than trial data. Based on the reliability analysis values of each scale, some items were removed with a view to maintaining the integrity of the scales.

Key interview candidates for data collection were determined to be primary stakeholders of the evaluations and/or members of the evaluation reference group, those responsible for undertaking the evaluations, and representatives of the organisations responsible for the delivery of the programs evaluated.

Prior to the time of the conduct of the interview, an information letter outlining the purpose of the study was given to stakeholders, and at the time of the interview they were asked to read and sign two copies of a Research Consent Form, one for them to keep and one for the investigator to keep on file. Copies of the Information Letter and the Research Consent Form are included at the front of Appendices A, B and C.

Data collection was undertaken by way of six interview schedules composed of both open response and rating scale items. Interviews were conducted with stakeholders and evaluators at a time prior to the conduct of the evaluation and following the release of the final evaluation report to stakeholders.

These interviews were conducted with stakeholders midway through the evaluation process to gather information regarding the influence of the evaluation process. Furthermore, information was gathered about the nature of the organizations sponsoring the program from representatives of the organisation not involved with the evaluated program. Organisational representatives were interviewed at the beginning of the evaluation study to gather information regarding characteristics of the organisation relevant to the organisational learning environment. Information was gathered in discussions with key stakeholders of the evaluations included as cases in this study, guided by identified issues of concern to this investigation, and in reviews of existing program literature (e.g. reports of program, program frameworks, and previous evaluations). Observations of stakeholders, as opportunity allowed, in attendance of meetings pertinent to the programs and evaluations, and in program delivery, have also been included as sources of information. Interview data was coded and entered into electronic data files. Data analysis, including reliability analysis and structural equation modelling and interpretation of the findings, immediately followed the completion of the final study interviews. A representation of the study procedure is presented in Table 4.5.

## DATA ANALYSIS

The scales of the instruments were analysed to determine their reliability and where appropriate, items were deleted from a scale to strengthen its reliability. The reliability value of the scale, however, was not the only consideration in determining whether to delete an item from a scale. Prior to deletion, items were also considered in relation to the content of other items in the scale as it was desirable to retain the range of item content to cover intended scale focus. The purpose of the scales was to provide measures for the observed variables indicating the latent variables of the Model of this study presented in Chapter III. The analysis of the items and scales is presented in Chapter V.

Structural equation modelling involves two phases. First, the measurement model is analysed. The measurement model represents the influence between the observed variables and the latent variables. The output of the measurement model analysis is the values representing the reliabilities and validities of these links. In the second phase of the structural equation modelling process the structural model is analysed. This involves testing the fit of the structure of the model to the data. In the instance of this study, structural equation modelling was undertaken of the eight versions of the Core Model of the study. Then further elaboration was undertaken of the Core Model in an exploratory analysis of versions of the Core Model, elaborated with additional latent variables. The results of the structural equation modelling process are presented in Chapter VII.

Qualitative data was gathered in open response items included in the interview schedules. Response categories were derived from data gathered in the pilot testing of the instruments.

### **Ethics**

This investigation was undertaken with the permission of the Murdoch University Human Ethics Committee and in accordance with the principles regarding the ethical and legal responsibilities of researchers toward their

subjects outlined in the National Health and Medical Research Council's 'Statement on Human Experimentation'. The study was conducted in accordance with the 'Rights of Human Subjects' standard as outlined by the Joint Committee on Standards for Educational Evaluation (1994).

## CONCLUSION

The purpose of this chapter is to provide details of the focus, design, procedures, timeline and instruments of this study. The evaluation studies included as cases in this study are described in Chapter VI. To some extent the characteristics of the cases included in this study have been determined by the criteria for inclusion in this study as detailed earlier in this chapter. However, some other organisation, program and stakeholder characteristics are distinctive to programs for which consent for inclusion was obtained. For instance, organisations willing to make a program evaluation and related documentation available for research are likely to be less protective of the work they are undertaking. Similarly, stakeholders working within organisations supportive of postgraduate research and willing to be interviewed regarding their part in the evaluation and the use of its information, are likely to hold similar values. Furthermore, the nature of the program evaluated had significant influence on whether or not it was available for investigation. Only programs that were open and public were made available for this study. Programs working with clients where sensitivity and confidentiality were critical (e.g. domestic abuse, elderly populations, etc) were excluded.

The instruments of the study are further discussed in Chapter V. The items of the interview schedule are considered in more detail and the reliability analysis of the scales are undertaken.





**Figure 4.2: Study Procedure**

## CHAPTER V

### DEVELOPMENT AND ANALYSIS OF INSTRUMENTS

The following six interview schedules of the study collected the qualitative and quantitative data of the investigation: Organisational Environment Interview Schedule, Stakeholder Interview Schedule I, Evaluator Interview Schedule I, Process Use Interview Schedule, Stakeholder Interview Schedule II, and the Evaluator Interview Schedule II. This chapter commences with an exploration of how the latent variables for the Structural Equation Model analysis were operationalised. It then considers each interview schedule in turn, including the logic underpinning the development of each instrument and a scale and item analysis. Each section, which is dedicated to an interview schedule, outlines the structure of the schedule, includes a description of the items in each scale, and a summarisation of the changes made as a result of the scale analysis process. The variables of the Intended Elaborated Model (refer to Figure 3.18) are drawn primarily from the scales in the various interview schedules. However, the *Program Theory* variable is based on classifications of program theory use in the evaluation studies and is described in Chapter VI

Obtaining evaluation case studies which met the criteria of this research study proved to be extremely difficult. This meant that an opportunity was not available to trial the scales on data from a separate, independent case study. Furthermore, even if another evaluation case study was available, its unique characteristics and limited number of observations may have provided misleading feedback information. Therefore, the analysis of the instruments undertaken was based on the actual data from the three case studies, rather than on trial data. Finally, more data was gathered in the conduct of the interviews than has actually been used to inform the findings of the study. This is the case because the opportunity to collect data was restricted to limited periods and therefore it was important to collect all the information expected, in order to reduce the likelihood of gaps in required information only emerging at a later stage.

## OPERATIONALISATION OF THE LATENT VARIABLES

The focus of this section is on the latent predictor variables and the outcome variables of the study. The structure of the variables is first considered followed by a description of the sources of data for the variables and the process by which data values were calculated, assigned or imputed. The final portion of this section summarises the scale analysis process.

### **LATENT VARIABLE STRUCTURE**

Each latent variable of the study is indicated by one observed variable. The data for the observed variables, with the exception of the Program Theory variable, comes from the scales included in the interview schedules. Some of the latent variables are indicated by only one observed variable. These include the *Process Use*, *Expected Study Involvement*, *Study Involvement (Post)*, *Commitment to the Study (Pre)*, and the *Commitment to the Study (Post)* predictor observed variables, and the *Influence of the Use of Program Theory in the Final Report*, *Influence of Involvement in Program Theory Elaboration*, and *Influence of Study Findings* outcome observed variables. The following latent variables by contrast are indicated by a number of observed variables scales each focusing on a particular facet of the latent variable: *Organisational Characteristics* (five observed variables), *Evaluator Characteristics* (five observed variables), *Stakeholder Characteristics* (four observed variables), and *Evaluation Study Characteristics* (four observed variables) predictor latent variables and the *Use of Evaluation Information* (four observed variables) outcome latent variable.

### **SOURCE OF DATA FOR THE OBSERVED VARIABLES**

All of the data for the observed variables measuring the *Organisational Environment Characteristics* latent variable was gathered through the Organisational Characteristics Interview Schedule. Data for the single Process Use observed variable measuring the *Process Use* latent variable, was gathered in the Process Use Interview Schedule, while the data for the

observed variables measuring the *Evaluator Characteristics* latent variable data gathering, was the focus of the Evaluator Interview Schedule I. Data for the observed variables measuring the *Stakeholder Characteristics*, *Expected Study Involvement*, and *Commitment to the Study (Pre)* latent variables was collected in the conduct of the Stakeholder Interview Schedule I. Finally, the Stakeholder Interview Schedule II was the source of the data for the observed variables measuring the *Evaluation Characteristics*, *Study Involvement (Post)*, and *Commitment to the Study (Post)* observed variables, as well as for the four outcome variables.

The interview group for the Organisational Characteristics Interview Schedule is distinct and separate from the interview groups for the other five interview schedules. Similarly, the evaluator interview group for the Evaluator Interview Schedules I and II is distinct from the interview groups for all of the other interview schedules. This means that the interview groups of these two interview schedules are entirely different to those from which the data for the outcome variables were collected. With regards to analysis of the data, the assessment of a relationship between two variables cannot be justified if the interview subjects are different for the two variables. The number of extraneous variables that might be responsible for any variation in responses between the two subject groups is impossible to control. To overcome this difficulty, an aggregated mean score, by program, was imputed and entered into the Stakeholder Interview II database for each observed variable indicating the Evaluator Characteristics and for the observed variables indicating the *Organisational Learning Environment Characteristics*.

Three groups were interviewed: organisational representatives, evaluators, and stakeholders, as explained in Chapter IV. As data on the outcome variables was collected only from stakeholders, and the stakeholders were grouped by program, it was necessary to assign a mean score for the observed variables Evaluator Characteristics and Organisational Learning Environment Characteristics. The data for each indicator variable for each interview group was used to calculate an aggregate score for each stakeholder interview subject who was included in the database to be subjected to quantitative analysis. This manner

of data reduction was adopted to simplify the structural model and yet still provide measures of the latent variables supported by detailed scale analysis (Bryne, 2001; Holmes-Smith, 2001).

It should be noted that although there is a substantial degree of overlap between the stakeholder groups interviewed for the Process Use Interview, Stakeholder Interview I and Stakeholder Interview II, it was not possible to include only the same respondents in each interview period. As there is considerable overlap between the subject groups of the Stakeholder I and Stakeholder II Interview Schedules, the data for those of the Stakeholder I subject group also interviewed in the second Stakeholder Interview, was introduced to the Stakeholder II data. The subjects interviewed for the Stakeholder II Interview but not the Stakeholder I Interview were given imputed values for the scales items, included in the Stakeholder Interview Schedule I. The imputed values were calculated as the mean of the aggregate item value by program. These values came from the scales indicating the *Stakeholder Characteristics*, *Expected Study Involvement* and *Commitment to the Study (Pre)* predictor latent variables.

The items developed to measure *Process Use*, as conceptualised for this study, were included in the Process Use Interview Schedule conducted midway through each evaluation study. Similar to the items measuring the *Stakeholder Characteristics* variable, data for Process Use subjects who also were interviewed for the Stakeholder II Interview Schedule, was introduced into the Stakeholder II Interview Schedule database, and values imputed for Stakeholder II Interview subjects who had not been interviewed for the Process Use Interview. The imputed values were the mean of the aggregate score calculated for each program, for the items measuring *Process Use*. The procedure for the assignment of the *Program Theory* variable value is described in Chapter VI.

## SCALE ANALYSIS PROCESS

The properties of each of the scales were analysed using the SPSS reliability analysis programme. The decision to retain or delete items from each scale was made based on the number of other scale items with which each item correlated ( $r_{ii} \geq 0.3$ ), the corrected item-total correlation ( $r_{it} \geq 0.3$ ) of each item, with Cronbach's Alpha  $\alpha \geq 0.7$  as the criterion for acceptance value and consideration of the item's contribution to the Alpha value. In addition, the concept that is the focus of an item and the extent to which this concept is accommodated by other items included in the scale is considered in terms of whether the item is needed in relation to the required content range of the scale.

Each scale item focuses on a distinct facet of the observed variable. There are likely to be situations where the items of a scale do not correlate at least moderately with each other, yet still have value with regards to gaining insight regarding the variable. At times, items with weak inter-item correlation values, or a weak item-total correlation value, that might otherwise have been deleted from a scale, have been retained for the contribution which the item makes to the range of the concepts on which the scale focuses.

## THE ORGANISATIONAL ENVIRONMENT INTERVIEW SCHEDULE

Pawson and Tilley (1997), in their book *Realistic Evaluation*, advocate that in the undertaking of an program evaluation, the relationship between the causal mechanisms of the program and their effects must be considered. However, they recognise that the relationship between the causal mechanisms and their effect is neither fixed nor definite. It is contingent on the presence of 'contextual conditions' to change the potential for a casual effect into a reality (p. 69).

In the present consideration of the use of evaluation information, the concept of 'contextual conditions' applies, in particular, the recognised importance of

the organisational environment, as a significant contextual influence on evaluation use (Jenlink, 1994; Mathison, 1994; Owen and Lambert, 1995; Preskill and Torres, 1999a, 2000a). The inclusion of the *Organisational Environment Characteristics* latent variable is important because if the organisation responsible for the delivery of a program has structures and practices in place which are prohibitive to the reception and use of evaluation information, the potential for use is diminished. Accounting for the organisation's environment enables an awareness of independent and extraneous organisational barriers that may influence evaluation use. The implications of the logic of organisational literature regarding organisational learning for evaluation practice are powerful. Simply, the greater the extent to which an organisation has a learning nature the more likely its members are to attend to a range of information, including evaluation information, in the development and conduct of the program they deliver, and the more likely they will be in a position to use that information in the decisions they make.

The concept of organisational learning has been in the literature for nearly thirty years (e.g. Argyris and Schön, 1978), but the concept of a learning organisation has more recently emerged (e.g. Senge, 1990; Watkins and Marsick, 1992). Learning organisations, as popularized by Peter Senge (1990), are places:

where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together (p. 1)

The work of Preskill and Torres considers the import of evaluative inquiry for organisational learning. They argue that for evaluative inquiry to be successful an organisation must have the infrastructure to support and facilitate it (1999a). The instrument included in this study to assess organisational characteristics is a modified version of the Readiness for Organisational Learning and Evaluation Instrument (ROLE) (2000b) for determining the extent to which a supportive infrastructure is present in an organisation. Theirs

is a diagnostic instrument developed to determine the presence of organisational aspects supportive of evaluative inquiry and organisational learning, including organisational culture, forms of communication, and systems and structures present to support evaluative inquiry.

The items of the Organisational Environment Characteristics Questionnaire included in this study also draw on the six requirements Watkins and Marsick (1992) advocate for strategies for the structuring of learning organisations. These are: 1) the creation of continuous learning opportunities; 2) the promotion of inquiry and dialogue; 3) the encouragement of collaboration and team learning; 4) the establishment of systems to enable shared learning; 5) the empowerment of organisation members in the development of a shared vision; and 6) the connection of the organisation to its environment.

The intent of this interview schedule is not to assess the extent to which organisational learning is occurring, but rather to determine the extent to which characteristics considered to be supportive of a learning environment are present in an organisation. The interview schedule includes five scales: *Perceived Characteristics of the Organisation Environment*; *Perceived Characteristics of Leadership Practice*; *Perceived Characteristics of Personal Practice*; *Characteristics of the Perceived Use of Information*; and *Perceived Characteristics of Teamwork*. Each of the five scales of this interview schedule are described below and the reliability analysis of each is detailed. The final section of the interview schedule is not a scale and seeks to gather additional pertinent information regarding the respondent.



## ORGANISATIONAL LEARNING ENVIRONMENT CHARACTERISTICS SCALE

Eleven items have been included in the interview schedule to gain insight into organisational representatives' perception of characteristics of the organisational environment. These eleven items were designed to form the *Perceived Characteristics of the Organisational Environment Scale*, which are the measure for *the Organisational Learning Environment Characteristics* observed variable of this study.

The items in the Organisation Environment section of the questionnaire are similar to those included in the sections addressing Leadership Practice and Personal Practice, although the wording of the item focusing on these factors has been tailored to the subject group of each section. For instance, the items of the section focusing on Organisation Environment refer to the 'organisation', Leadership Practice items refer to 'managerial staff' and Personal Practice items refer to 'you'. The final section includes a few questions to gather individual and demographic information, with a view to adding insight to responses to the scale items.

The items are designed to gather information representative of a complex characteristic which one manager described as 'the smell of the place' made apparent in the multitude of details on organisational behaviour and operation. As Bartlett and Ghoshal (1995) write, "It is as pervasive and influential as climate - just as one can be energized by the fresh, crisp air at the mountain resort in spring, so too can the behavioural context of a company provide people with a source of stimulation." (p.12).

The thinking supporting the inclusion and focus of this factor rests upon the behavioural perspective, which considers the role of environment as a determinant of human behaviour. Simply, if the environment of the organisation is genuinely supportive of learning practices, the individuals of the organisation are more likely to pursue learning activities and seek new knowledge. Social unity and organisational cohesiveness are prized (Bartlett

and Ghoshal 1995). Barriers created by rigid autocratic leadership styles and an imposed hierarchy can be seen to encourage an atmosphere of fear, distrust, interrupted communication and fragmented work practices, and obstructed learning (Dilworth 1995). Luthans (1998) argues that job stress may be reduced through the creation of a more supportive environment in which employees are more likely to feel they have some control over their jobs. However, he notes that empirical support for this link is not yet available.

In the section focusing on the environment of the organisation, items 1 through 9 are intended to gather information regarding the extent to which the organisational environment is supportive or inhibitive of staff learning. The final two items, Item 10 and Item 11, target information regarding the extent to which the organisation is supportive or inhibitive of the trialling of new knowledge. The eleven items were included to form the *Organisation Environment Characteristics Scale*.

The first item of the section focuses on the extent to which there is a clear understanding of the organisation's vision, or what it is seeking to achieve (Item 1). A clear and common understanding of the organisation's intention(s) enables staff to consider their work in light of the organisational vision and in terms of the broader organisational goals.

It could be argued that the presence of a stressful environment makes learning difficult. If members of staff are over-worked they are less likely to undertake learning activities. In fact, with little time to critically reflect on their regular work tasks they are likely to overlook opportunities to learn (Preskill and Torres 1999a). Lack of insight may lead to a misplacement and loss of resources. Item 2 has been included to gain insight into organisational representatives' perceptions of expected staff workload.

An organisation that has realistic expectations in terms of staff workload indicates a balanced awareness of staff, their capacities and skills, and realistic goals. Members of staff who are guided by this balanced realism may be more trusting of the environments in which they work. Individuals operating in an

organisation where they find they face unachievable or unrealistic workload expectations may be more likely to feel threatened and react to this threat in ways that will increase defensiveness and reduce the probability of learning to learn (Argyris 1999). Watkins and Marsick (1992) advocate action and reflection as the two necessary components of a learning process enabling learning from experience as opposed to simply having an experience. Items have been included in this section to gain insight into the interest and support for learning new knowledge and skills (Item 4) and the proactive intent towards the improvement of policies, programs or services (Item 3).

Three items (items 5, 6, and 7) have been included to gain insight into the transparency of the organisation in terms of encouragement regarding negative/constructive feedback relating to organisation practices, policies, assumptions, and so on. The level of transparency evident in an organisation and in the work practices of staff, is a reflection of trust inherent in the organisational environment and an understanding of common goals. Trusting relationships are built incrementally through a number of practices guided by consistency, honesty, integrity and openness (Preskill and Torres 1999b).

Competition, a driving force in the traditional organisation environment, tends to inhibit trust among organisation staff and leaders, in turn impeding free dialogue and information exchange. Competition within the organisation is a barrier to learning. “The power of the behavioural context lies in its impact on the behaviour of individual organisation members” (Bartlett and Ghoshal 1995, p. 18). The encouragement of competition among organisational branches or units (Item 8) latent in the environment of the organisation reduces opportunity for collaboration.

Item 8 has been included to gain insight into the element of competition embedded in the work practices of the organisation. Alternatively, co-operation encourages trust, a practice that encourages information exchange and dialogue among employees. Item 9 has been included to gain insight into the extent of co-operative work practice supported, by and in, the organisation.

Item 10 focuses on the characteristic of support for trialling of new knowledge or skills. Item 11 also focuses on support for trialling of new knowledge or skills and for the risk-taking element necessary to implement the trials. If there is little or no perceived support for the use of new knowledge and skills in the trialling of new programs, policies, strategies, and so on, then it is unlikely staff will create or undertake such opportunities. If opportunities to apply new knowledge and skills are not present, learning is truncated as experimentation with new approaches is an important step in the learning strategy (Garvin 1993). Furthermore, if staff are aware of alternative practices possibly superior to present organisational routines, but are unable to implement changes, they are likely to become discouraged from undertaking further efforts to learn.

All items, with the exception of Item 11, require a response on a five point scale. Item 11, focusing on organisation support of risk-taking, requires respondents to choose between two options regarding the organisation's reaction to unsuccessful trialling of policies, strategies, programs, and so on, then asks them to rate on the five point scale the strength of that choice as an organisational preponderance. In the item analysis, Item 11 has been modified to include only positive ratings regarding the organisation's reaction to unsuccessful trialling of policies, strategies, programs, and so on. All negative ratings have been scored as "0".

#### ORGANISATIONAL LEARNING ENVIRONMENT: CHARACTERISTICS OF THE SCALE ITEM ANALYSIS

The scale analysis first calculated a Cronbach's Alpha which was  $\alpha=0.89$ , but with the deletion of Item 8 the scale Alpha was increased to  $\alpha=0.92$ . Item 8 was deleted from the scale due to low inter-item correlations with other scale items ( $r_{ij} \geq 0.3$  with only 1 other scale item) and low item-total correlation ( $r_{it} = 0.03$ ). Furthermore, the characteristic focus of the item on competitiveness is also a reverse focus of Item 9 addressing collaboration. Inter-item correlations for the items ranges from  $r_{ij} = -0.21$  to  $r_{ij} = 0.82$ .

Although Item 1 has low inter-item correlations ( $r_{ii} \geq 0.3$  with only four items) and a low item-total correlation ( $r_{it} = 0.22$ ), the deletion of the item from the scale would only have increased the value of Cronbach's Alpha to  $\alpha = 0.93$ . Therefore, a decision was made to keep the item in the scale for the range it adds to the scale focus, as it is the only scale item concerned with the organisation's vision or aims. The other scale items correlate ( $r_{ii} \geq 0.3$ ) with six items or more. The final scale of ten items has a mean of 39.25 and a standard deviation of s.d. = 9.46. A summary of reliability analysis is presented in Table 5.1.

### **PERCEIVED LEADERSHIP PRACTICE SCALE**

Twelve items have been included in the interview schedule to gain insight into organisational representatives' perception of leadership practice characteristics relevant to organisational learning of those employed in a managerial or supervisory capacity by the organisation. These twelve items were intended to form the *Perceived Characteristics of Leadership Practice Scale* of this study, which provides the measure the *Leadership Practice* observed variable.

As Preskill and Torres (1999b) write, "Leadership is not just telling people what is important and what to do. It is about consistently modelling the behaviour that leaders wish to see in their employees" (p. 164).

The items included in the Organisational Environment section for this section of the questionnaire have been reworded to reflect the focus on leadership practice. Only one additional item (Item 19) has been added to gain insight into the respondents' perception of the transparency of those in a managerial or supervisory capacity, in terms of their openness to offering constructive feedback regarding the policies or practices of the organisation. Item 12 to Item 21 inclusive gather information regarding the extent to which the managerial staff of the organisation model, and are supportive of, staff learning. The final two items, Item 22 and Item 23, collect information regarding the extent to which the managerial staff of the organisation is

supportive or inhibitive of trialling of new knowledge. The twelve items were included to form the perceived characteristics of the *Leadership Practice of the Organisation Scale*.

**Table 5.1: Perceived Characteristics of the Environment of the Organisation Scale. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{it} \geq 0.3^1$	$r_{it}^2$
<b>Of aims or vision of the organisation.</b>			
ITEM 1	Is there a clear understanding of what the organisation is seeking to achieve?	3	0.22
<b>Of the organisation's expectation of staff workloads.</b>			
ITEM 2	Are the expectations of the organisation regarding staff workload realistic?	6	0.75
<b>Of the proactive nature of the organisation regarding improvement of policies, practices, strategies, etc</b>			
ITEM 3	Does the organisation regularly look for ways to improve policies, programs or services?	6	0.41
ITEM 4	Does the organisation support the learning of new knowledge and skills by the staff?	9	0.71
<b>Of organisation transparency regarding policies or practices.</b>			
ITEM 5	Is open dialogue regarding organisation policies or practices encouraged?	7	0.78
ITEM 6	Is negative feedback regarding the policies or practices of the organisation encouraged?	7	0.79
ITEM 7	Are all assumptions, policies or practices open for discussion?	8	0.84
<b>Of encouragement of competitiveness/collaboration within the organisation.</b>			
ITEM 8	Is competition encouraged among the branches or units of the organisation?	1 <sup>3</sup>	
<b>Item deleted due to low <math>r_{it}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.03</math>). When item deleted <math>\alpha = 0.92</math>.</b>			
ITEM 9	Is collaboration among the organisation's branches or units encouraged?	8	0.76
<b>Of organisational support for trialling of new knowledge and skills- or risk-taking.</b>			
ITEM 10	Does the organisation support the trialling of new knowledge and skills by the staff?	9	0.90
ITEM 11 <sup>4</sup>	To what extent are unsuccessful undertakings considered in terms of knowledge gained from the experience?	9	0.86
Note:			
<sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{it} \geq 0.3$ ).			
<sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.			
<sup>3</sup> Represents the number of other scale items the item correlates with ( $r_{it} \geq 0.3$ ) in first run of scale with all items included.			
<sup>4</sup> Item reworded for presentation in table.			

All items, with the exception of Item 23, require a response on a five point scale, one representing 'Not at All' and five representing 'A Great Deal'. Item 23 focusing on managerial support of risk-taking, requires respondents to choose between two options regarding the reaction of managerial staff to unsuccessful trialling of policies, strategies, programs, and so on, then asks them to rate on the five point scale the strength of that choice as a preponderance of managerial staff. In the item analysis, Item 23 has been modified to include only positive ratings regarding the reaction of managerial staff to unsuccessful trialling of policies, strategies, programs, and so on. All negative ratings have been scored as "0". The reliability analysis of the scale item analysis is summarised in the following section.

#### Perceived Leadership Practice: Scale Item Analysis

Data analysis began with the inclusion of the 12 items listed in the table below to gather information regarding the organisational representatives' perceptions of managerial staff or leadership practice. The first Cronbach's Alpha calculated was  $\alpha = 0.86$ , but with the deletion of Item 12 in the reliability analysis, the scale Alpha was increased to  $\alpha = 0.88$ . The decision to delete Item 12 was made due to weak correlations with other scale items, ( $r_{ii} \geq 0.3$  with one item), and a very low item-total value ( $r_{it} = 0.051$ ). Similarly, Item 14 was deleted from the scale due to low inter-item correlation values ( $r_{ii} \geq 0.3$  with three items) and a very low item-total value ( $r_{it} = 0.21$ ). The deletion of Item 14 increased the scale Alpha value to  $\alpha = 0.89$ . Furthermore, the range of the scale was not significantly compromised as Item 15 is of a similar focus to Item 14. Finally, although Item 18 correlates with only four other final scale items ( $r_{ii} \geq 0.3$ ), it has an acceptable item-total correlation value of  $r_{it} = 0.39$ . Furthermore, the deletion of Item 18 would only have increased Cronbach's Alpha value to  $\alpha = 0.90$ . Therefore, the item was retained for the range it adds to the scale.

Inter-item correlations for the final scale items range from 0.05 to 0.84. All final scale items correlate with four or more items ( $r_{ii} \geq 0.3$ ). The final scale of ten items has a mean of 36.13 and a standard deviation of s.d.= 8.09. A summary of reliability analysis is presented in Table 5.2.

**Table 5.2: Perceived Characteristics of the Leadership Practice of the Organisation Scale. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Clarity regarding the vision or aims of the organisation among managerial staff.</b>			
ITEM 12	Do managerial staff have a clear understanding of what the organisation is seeking to achieve? <b>Item (1<sup>st</sup>) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.05</math>). When item deleted <math>\alpha = 0.88</math>.</b>	1 <sup>3</sup>	
<b>Managerial staff's management of work commitments.</b>			
ITEM 13	Do managerial staff make realistic work commitments for themselves?	7	0.75
<b>Managerial staff proactive nature regarding improvement of policies, practices, strategies, etc...</b>			
ITEM 14	Do managerial staff regularly look for ways to improve policies, processes, programs, or services? <b>Item (2<sup>nd</sup>) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.21</math>). When item deleted <math>\alpha = 0.89</math>.</b>	4 <sup>3</sup>	0.25
ITEM 15	Do managerial staff model the importance of learning through their own efforts to learn?	9	0.84
<b>Managerial staff transparency regarding policies or practices.</b>			
ITEM 16	Do managerial staff openly talk about organisation policies or practices?	7	0.61
ITEM 17	Do they offer constructive feedback regarding the policies or practices of the organisation?	7	0.78
ITEM 18	Are there any assumption, practices, policies, etc., that managers refrain from offering an opinion or viewpoint on?	4	0.39
ITEM 19	Are managerial staff open to constructive feedback from staff?	8	0.60
<b>Managerial staff encouragement of staff collaboration within the organisation.</b>			
ITEM 20	Do managerial staff encourage staff to work together?	8	0.79
ITEM 21	Do managerial staff encourage those they supervise to compete with each other?	5	0.42
<b>Managerial support for trialling of knowledge and skills- or risk-taking.</b>			
ITEM 22	Are managerial staff supportive of trial programs, policies and strategies based on new knowledge?	7	0.78
ITEM 23 <sup>4</sup>	To what extent do managerial staff consider unsuccessful organisational policies, strategies, programs, etc., in terms of knowledge gained from the	8	0.79
<p><b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with (<math>r_{ii'} \geq 0.3</math>).  <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.  <sup>3</sup> Represents the number of other scale items the item correlates with (<math>r_{ii'} \geq 0.3</math>) in first run of scale with all items included.  <sup>4</sup> Item reworded for presentation in table.</p>			



## **Perceived Characteristics of Personal Practice Scale**

Twelve items have been included in the interview schedule to gain insight into organisational representatives' perception regarding characteristics of their personal practice relevant to organisational learning. These twelve items were intended to form the *Perceived Characteristics of Personal Practice Scale*, proving the measure for the Personal Practice observed variable of this study.

The importance of individual learning for organisational learning is both subtle and obvious: obvious because all organisations are composed of individuals and subtle because organisations can learn independent of any specific individual but not independent of all individuals (Kim 1993). Similarly, Jenlink (1994) and Senge (1990, 1994) identify organisational learning as the intersection of personal knowledge of the member and social knowledge of the organisation.

Argyris and Schön (1978) consider the contribution of the individual in their hierarchy of learning levels regarding organisational learning. The first level of their hierarchy, *single-loop learning*, is learning that leads to the development and clarification of existing knowledge structures. The second level, *double-loop learning*, occurs when new knowledge challenges existing knowledge structures. New forms of knowledge structure are conceptualised and then, in turn, refined and developed in reference to new knowledge through the lens of reconceptualised knowledge structures- or ways of viewing the information presented to them. Finally, the third level of learning, *deutro-learning*, is the capacity to learn how to learn. As the individuals in an organisation learn how to learn, they create "...new organisational learning systems and a new culture that sanctions such learning" (Argyris, Putnam and Smith 1985, p. 153).

The items of this section are similar to those included in the previous two sections of the questionnaire, with changes made to item wording necessary to focus on the personal practice of the respondent, rather than perceptions of the organisation or the practices of managerial staff. Item 24 to Item 33 inclusive

focus on the extent to which the personal practice of the respondent is supportive of organisational learning. The final two items, Item 34 and Item 35, target information regarding the extent to which the personal practice of respondents is supportive or inhibitive of trialling of new knowledge.

All items, with the exception of Item 35, require a response on a five point scale. Item 35, focusing on the personal perception of stakeholder's regarding risk-taking, requires respondents to choose between two options regarding their reaction to unsuccessful trialling of policies, strategies, programs, and so on, then asks them to rate on the five point scale the strength of that choice as a preponderance. In the item analysis Item 35 has been modified to include only positive ratings regarding the personal reaction of stakeholders to the unsuccessful trialling of policies, strategies, programs, and so on. All negative ratings have been scored as '0'.

Certainly, the respondent may be a manager, in which case he or she may answer the items of both sections (Leadership and Personal Practice) in consideration of their own practice. However, the items of the Leadership Practice section are designed to more generally address the practice of the managerial group and not personal use. It is expected that managers answering the questions will comment with regards to general organisational managerial practice and with regards to personal practice in the Leadership items of this section.

#### PERCEIVED CHARACTERISTICS OF PERSONAL PRACTICE: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the 12 items listed in the Table 5.3 below to gather information from organisational representatives regarding their personal practice. Cronbach's Alpha of all 12 items was  $\alpha = 0.62$ , but with the deletion of four items (Item 26, Item 27, Item 32, and Item 33) in the reliability analysis, the scale Alpha was increased to  $\alpha = 0.74$ . In the case of all four items deleted, the inter-item ( $r_{ii'} \geq 0.3$  with two or fewer items) and item-total correlation values were weak (Item 26  $r_{it} = -0.01$ , Item 27  $r_{it} = -0.22$ , Item 32  $r_{it} = 0.22$ , Item 33  $r_{it} = 0.08$ ). The decisions to delete Item 26 and Item

27 from the scale were difficult to make as both items focus on characteristics which no other scale item deals with. Item 26 considers the proactive nature of the respondent in terms of the extent to which they look for opportunities to improve policies, processes, programs or services, while Item 27 deals with the extent to which respondents pass up opportunities to learn new skills and knowledge. The deletion of these items narrows the intended concepts measured by the scale, but the low inter-item correlations of these items with other scale items indicates they do not relate with the other scale items. Similarly, the decision to delete both items 32 and 33 focusing on the competitiveness/co-operation of the respondent over work issues was also difficult, as these are the only two scale items focusing on this area. However, the results of the reliability analysis indicate they are not a good fit with the scale. Therefore, in the interest of maintaining a strong and useful scale, the items have been deleted.

Inter-item correlations for the final scale items ranges from  $r_{ii'} = -0.08$  to  $r_{ii'} = 0.76$ . Item 35 has a  $r_{ii'} \geq 0.3$  with two other final scale items. The other items have a  $r_{ii'} \geq 0.3$  with three items or more scale items. The final scale of nine items has a mean of 30.98 and a standard deviation of s.d.= 5.30. A summary of reliability analysis is presented in Table 5.3.

#### **PERCEIVED CHARACTERISTICS OF USE OF INFORMATION SCALE**

Twelve items have been included in the interview schedule to gain insight into organisational representatives' perception of characteristics relevant to the organisation's use of information. These 12 items were intended to form the perceived characteristics of the organisation's *Use of Information Scale* providing the measure for the observed variable organisational Use of Information of this study. The reliability analysis of the scale item analysis is summarised in the following section.

Rich's (1991) conceptual paper illuminates his perspectives regarding knowledge creation, diffusion and utilisation, recognises indicators of the

stability and continuity of information flow over time in the determination of organisational information use. His recognition of the influence of organisational rules and practices in the constraint of information exploration, gathering and selection activities, is pertinent to the work proposed here. In his conceptual article reviewing successful and unsuccessful implementation of learning practices in organisations, Garvin (1993) suggests organisations with a commitment to learning are good at transferring knowledge quickly and efficiently throughout the organisation, relying on established practices, not short-lived, spurious nor episodic practices. The items of this section have been adapted from the Communication of Information and Evaluation sections of the Readiness for Organisational Learning and Evaluation (ROLE) (Preskill and Torres 2000b) instrument. Items 36, 37, 38, 39 and 40 focus the information collection routines of the organisation. Items 41, 42, 46 and 47 are concerned with information dissemination practices of the organisation. Finally, items 43, 44 and 45 have been included to gather information regarding the use of information collected by the organisation by staff.

#### PERCEIVED CHARACTERISTICS OF THE ORGANISATION'S USE OF INFORMATION: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the 12 items listed in Table 5.4 below to determine stakeholder perceived characteristics regarding the organisation's use of information. Cronbach's Alpha of all 12 items was  $\alpha = 0.93$ . With the deletion of Item 42, due to weak inter-item correlations with other scale items ( $r_{ij} \geq 0.3$  with two other scale items) and a low inter-item correlation values ( $r_{ii} = 0.25$ ), Cronbach's Alpha increased to  $\alpha = 0.95$ .

Inter-item correlations for the final scale items range from  $r_{ii} = 0.22$  to  $r_{ii} = 0.82$ . All final scale items correlate ( $r_{ij} \geq 0.3$ ) with eight or more other scale items. The final scale of 11 items has a mean of 38.31 and a standard deviation of s.d. = 9.99. A summary of the reliability analysis is shown in Table 5.4.

**Table 5.3: Perceived Characteristics of the Personal Practice Scale. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii} \geq 0.3^1$	$r_{ii}^2$
<b>Perception of personal clarity regarding the vision or aims of the organisation.</b>			
ITEM 24	Do you have a clear understanding of what the organisation is seeking to achieve?	3	0.40
<b>Perception of personal management of work commitments.</b>			
ITEM 25	Do you make realistic work commitments for yourself?	3	0.47
<b>Perception of personal proactivity regarding improvement of policies, practices, strategies, etc...</b>			
ITEM 26	Do you regularly look for ways to improve policies, processes, programs, or services?		
	<b>Item (2<sup>nd</sup>) deleted due to low <math>r_{ii}</math> values. Item-total value (corrected) of item low (<math>r_{ii} = -0.01</math>). When item deleted <math>\alpha = 0.70</math>.</b>	2 <sup>3</sup>	
<b>Perception of personal effort to learn new knowledge and skills.</b>			
ITEM 27	Do you pass up opportunities to learn new knowledge and skills?		
	<b>Item (1<sup>st</sup>) deleted due to low <math>r_{ii}</math> values. Item-total value (corrected) of item low (<math>r_{ii} = -0.22</math>). When item deleted <math>\alpha = 0.68</math>.</b>	0 <sup>3</sup>	
<b>Perception of personal transparency regarding policies or practices.</b>			
ITEM 28	Do you openly talk about organisation policies or practices?	4	0.54
ITEM 29	Do you give negative feedback regarding organisation policies or practices?	5	0.50
ITEM 30	Are there any assumption, practices, policies, etc., that you	3	0.42
ITEM 31	Do you encourage negative feedback from your colleagues at work?	5	0.58
<b>Perception of collaboration/Competitiveness in personal work practice.</b>			
ITEM 32	Do you co-operate with your workmates over work issues?		
	<b>Item (4<sup>th</sup>) deleted due to low <math>r_{ii}</math> values. Item-total value (corrected) of item low (<math>r_{ii} = 0.22</math>). When item deleted <math>\alpha = 0.75</math>.</b>	2 <sup>3</sup>	
ITEM 33	Are you competitive with your workmates over work issues?		
	<b>Item (3<sup>rd</sup>) deleted due to low <math>r_{ii}</math> values. Item-total value (corrected) of item low (<math>r_{ii} = 0.08</math>). When item deleted <math>\alpha = 0.74</math>.</b>	2 <sup>3</sup>	
<b>Perception of personal uptake of opportunities to trial new knowledge and skills- or risk-taking.</b>			
ITEM 34	Do you take opportunities to trial programs, policies or strategies based on new knowledge?		
		3	0.56
ITEM 35	To what extent do you consider unsuccessful organisational policies, strategies, programs, etc., in terms of knowledge gained from the experience?		
		2	0.32
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents the number of other scale items the item correlates with ( $r_{ii} \geq 0.3$ ) in first run of scale with all items included. <sup>4</sup> Item reworded for presentation in table.			

## PERCEIVED CHARACTERISTICS OF TEAMWORK SCALE

Nine items have been included in the interview schedule to gain insight into organisational representatives' perception of characteristics relevant to teamwork undertaken in the organisation. These nine items were intended to form the *Perceived Characteristics of the Organisation Teamwork Scale* of this study, proving the measure for the Teamwork observed variable. The reliability analysis of the scale item analysis is summarised in the following section.

The presence of these characteristics supports the development of learning groups with a shared understanding of the path to job completion with commonalities among the way members talk and think about their work. It is in these groups that the most valuable and innovative learning takes place. Teamwork lessens the merits of individual performance and competitiveness, encouraging team or unit members to focus on the collective impact of their work, thereby surpassing the gains achievable by individuals working alone. Furthermore, team members are then able to share what is learned in the team environment throughout the organisation as they return to their work outside the team, again maximising the flow of information to other organisations' members and the learning of the organisation (Preskill and Torres, 1999b).

The items of this section have been adapted from the Teams section of the Readiness for Organisational Learning and Evaluation (ROLE) (Preskill and Torres 2000b) instrument. If respondents do not have opportunities to work as part of a team (Item 48) they are not asked to respond to this section. Items 49 to 57 inclusive focus on a range of teamwork characteristics including: the organisation's support of team work in the provision of training regarding how to work as a team member and team meeting attendance (items 49 and 50); the presence of open, task-oriented dialogue (Item 51); the development of a focus or vision (Item 52); conflict resolution (items 53 and 54); the productivity of work undertaken in teams (items 55 and 56); and knowledge transference from the team to other members of the organisation (Item 57).

**Table 5.4: Perceived Characteristics of the Organisation's Use of Information. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Information collection routines of the organisation.</b>			
ITEM 36	Is information regularly gathered from clients and other stakeholders to gauge how well programs and activities are doing?	10	0.76
ITEM 37	Are there adequate records of past change efforts and what happened as a result?	10	0.84
ITEM 40	Does the organisation collect information it needs?	10	0.73
<b>Information collection/evaluation routines of the organisation.</b>			
ITEM 38	Are formal evaluations of the organisation's activities and programs routinely undertaken?	9	0.64
ITEM 39	Are formalised evaluation activities incorporated into the delivery of organisational programs?	9	0.59
<b>Information access/ dissemination routines of the organisation.</b>			
ITEM 41	Are systems adequate to disseminate information gathered by the organisation to those staff who need and can use it?	10	0.78
ITEM 42	Is it difficult for staff to access information collected by the organisation to make decisions regarding their work? <b>Item deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.27</math>). When item deleted <math>\alpha = 0.95</math>.</b>	2 <sup>3</sup>	0.27
ITEM 46	To what extent does the organisation alert staff to new information?	9	0.93
ITEM 47	Does the organisation encourage staff to share information?	9	0.70
<b>Information use routines of the organisation staff.</b>			
ITEM 43	Do staff use the information collected by the organisation in their work practice?	10	0.80
ITEM 44	Do staff rely on the information collected by the organisation in the decisions they make regarding their work?	9	0.70
ITEM 45	Do managerial staff tend to use information collected by the organisation in their decisions?	10	0.88
Note:	<sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ) in first run of scale with all items included.		

#### PERCEIVED CHARACTERISTICS OF ORGANISATION TEAMWORK SCALE: SCALE ITEM ANALYSIS.

Data analysis began with the inclusion of the nine items listed in the Table 5.5 below to determine stakeholder commitment to the program. Cronbach's Alpha ( $\alpha$ ) of all nine items was  $\alpha = 0.78$ , but with the deletion of Item 56 in the reliability analysis, the scale Alpha was increased to  $\alpha = 0.81$ . Item 56 focusing on the effectiveness of teamwork was deleted from the scale at an early point due to low inter-item correlations with other scale items ( $r_{ii'} \geq 0.3$  with only 1 item) and a low item-total correlation value ( $r_{it} = -0.09$ ). As Item 55 also deals with the effectiveness of work undertaken by teams, the characteristic is included in the scale.

Inter-item correlations for scale items range from 0.08 to 0.80. Item 49 has an  $r_{ii'} \geq 0.3$  with three other scale items. It has been kept in the scale in spite of low inter-item correlation values with other items, as it is the only item focusing on the organisation's provision of support for teamwork through the provision of training, and it has a reasonable item-total correlation value ( $r_{it} = 0.39$ ). Furthermore, the deletion of Item 49 from the scale would have increased Cronbach's Alpha only marginally to  $\alpha = 0.83$ .

Item 55 correlates ( $r_{ii'} \geq 0.3$ ) with four other final scale items. The other items of the scale have an  $r_{ii'} \geq 0.3$  with five items or more. The final scale of eight items has a mean of 29.00 and a standard deviation of s.d. = 6.02. A summary of the reliability process is presented in Table 5.5.

#### THE FINAL SECTION

The final section of the interview schedule includes items focused on gathering demographic information regarding the organisational representatives, with a view to providing insight to their interview responses. The items focus on their position at the organisation (Item 58) and the number of years they have been employed in the position (Item 59), the number of



years they have worked for the organisation (Item 60) and how long they intend to stay with the organisation (Item 61).

## **SUMMARY**

The intent of this questionnaire is not to assess the extent to which organisational learning is occurring, but rather to determine the extent to which characteristics considered to be supportive of a learning environment in an organisation are present. As Schein (1992) writes, “The learning culture must be built on the assumption that communication and information are central to organisational well-being” (p. 370).

The focus of the above text has been on the items included in the six sections of the Organisational Interview Schedule. The logic underpinning the sections and the items contributing to the five scales included in them, has been explained. It is believed, for the purposes of this paper, that learning is a cyclical phenomenon, rather than linear, in which a number of processes interact. The challenge for this questionnaire has been to develop items that sample activities, processes, and undertakings that represent the full learning cycle and finish with a questionnaire of reasonable length. The items have been ordered in a logical fashion to maximise the flow of item presentation, rather than in accordance with any priority.

The five scales detailed above represent the measures for the five observed variables indicating the Organisational Environment Characteristics latent variable included in this study expected to have influence on the use of the information of the evaluations included in this study. Interview data was used for analysis. With regards to each scale, a few items with weak inter-item correlation ( $r_{ii} \leq 0.03$ ) have been deleted. All study scales have been verified (Cronbach's Alpha  $\geq 0.7$ ). The analysis of the scales helped confirm what was anticipated in terms of scale and sub-scale content and focus.

**Table 5.5: Perceived Characteristics of the Organisation's Teamwork Scale. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Organisation support of teamwork through provision of training to ensure effective teamwork.</b>			
ITEM 49	Are employees provided training on how to work as a team member?	3	0.40
<b>Team meeting attendance reflects both team effectiveness and organisation support.</b>			
ITEM 50	Are team meetings well attended?	5	0.67
<b>Open communication and development of shared vision as part of teamwork- democracy.</b>			
ITEM 51	Is task-oriented dialogue a part of team meetings?	5	0.62
ITEM 52	Is an effort made at team meetings to develop a vision of the team tasks and goals that are <i>shared</i> by all team members?	5	0.57
ITEM 53	Is conflict that arises among team members resolved effectively?	6	0.71
ITEM 54	Do team members go along with decisions they don't really agree with?	6	0.47
<b>Effectiveness of teamwork.</b>			
ITEM 55	Are teams more productive than individuals working alone?	4	0.40
ITEM 56	Do teams fail to accomplish work they are charged to do? <b>Item deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = -0.09</math>). When item deleted <math>\alpha = 0.81</math>.</b>	1 <sup>3</sup>	
<b>Effectiveness of information dissemination from team to organisation.</b>			
ITEM 57	Is knowledge developed in teams shared with other groups or co-workers?	6	0.58
Note: <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ) in first run of scale with all items included.			

## STAKEHOLDER INTERVIEW SCHEDULE I

A stakeholder, as stated by Scriven (1991), is a person who has a substantial commitment to a program in terms of "...ego, credibility, power, futures or other capital invested in the program, and thus can be held to be to some degree at risk with it" (p. 334). Characteristics of program stakeholders have been found by numerous authors to have influence on the use of the evaluation information (e.g. Cousins and Walker 2000, Cummings 1997, Hudson-Mabbs 1993, Leviton and Hughes, 1981, Vlahov 1990). In the case of this study, all stakeholders held positions in which some or all of their work time was committed to the program.

The purpose of the Stakeholder Interview Schedule I was to gather information from primary stakeholders of each evaluation study included in this study. The interview schedule includes six scales: *Stakeholder Commitment to the Program*; *Stakeholder Commitment to the Evaluation Study*; *Stakeholder Expectation of Involvement in the Evaluation Study*; *Stakeholder Characteristics- Education and Training in Social Research and Evaluation*; *Opportunity of Stakeholders to Use Information*; and, *Stakeholder Commitment to the Organisation*. The description and reliability analysis of each of the six scales of this interview schedule is detailed in the following sections.

All items of the interview schedule, with the exception of items 1, 2, 14, 16, 21, 23, 24, 37, 39, 40, 41, 46, 48, 49 and 50 require a response on a five-point scale. The recoding of Item 2, Item 14, and Item 21 to fit with the scale analysis is detailed below. The remaining items are all open response items necessary for gathering qualitative information to provide additional detail to the findings based solely on the quantitative data of the interview schedule scale items.

## PERCEIVED STAKEHOLDER COMMITMENT TO THE PROGRAM SCALE

The key finding of Patton et al. (1977) was the identification of the ‘personal factor’, “...made up of equal parts of leadership, interest, enthusiasm, determination, commitment, aggressiveness and caring..” (p. 73) of individual people. They found that when the personal factor emerged, evaluation had an impact. When it did not, impact was absent. Thirteen items (items 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14/15) have been included in the interview schedule to gain insight into stakeholder perception of characteristics of their commitment to the program. These 13 items were intended to form the *Stakeholder Commitment to the Program Scale* of this study. The reliability analysis of the scale item analysis is summarised in the following section.

Items 2, 6, 9, 10, 14 and 15 have been adopted from an instrument developed by Cummings (1997) to determine stakeholder commitment to the program. Although Cummings’ work included only one item to assess stakeholder involvement with the program, the five items included here (Item 1, Item 2, Item 3, Item 4 and Item 5) address involvement in both early and current program development and delivery.

Items 1 and 2 are the only open response items in this section. The first two items of the interview focus on how the interviewee became involved with the evaluation and the role they played in the undertaking of the evaluation. Item 1 focuses on the number of years stakeholders have been involved with the program as an indicator of commitment. The logic underpinning this item is that the longer a stakeholder has been involved with a program, the greater their commitment to that program.

Item 2, focusing on how stakeholders have been involved with the program, required the development of a rating scale regarding the role of the stakeholder in relation to the program. Stakeholders were selected for inclusion in the interviews in accordance with the role they held in relation to the program. The range of pertinent stakeholder roles varied slightly between programs, however, all roles generally fell into one of nine categories. To determine a

rating of each role, an instrument listing the roles was send electronically to a range of colleagues with a background in program evaluation. This instrument has been included in Appendix D. They were asked to rank the roles from 1 to 9 (1 being the stakeholder role with the highest stake in the program). No two roles could be given the same ranking. The overall ranking for each role was then determined by averaging the ranks given by each colleague. This rank for each stakeholder in turn was included as one of the factors considered influential to the stakeholder commitment to the program, the focus of the first 15 items of the Stakeholder Interview Schedule I.

Those with no direct stake in the evaluated program (organisation representatives) were given a '0' in the first column and a value of '99' (the code for not applicable) for the remaining four columns. Those with only one duty with regard to the program were coded for that duty. The code values incorporating the '0' have also been included in Appendix D.

In recoding the variable it was decided that a strategy had to be devised to represent the interviewee stake in the program in a more efficient manner. To do this, the number of roles the interviewee played in relation to the program has been added to represent a single value. Each role has been given one point, and the number of roles simply added to obtain a final score of program stake. The role of program Manager/ Co-ordinator was exempted from this process.

The role of Program Manager/ Co-ordinator is one in which the subject is involved in a more intricate basis with the program and as such, has a more detailed understanding of the program processes. The time and responsibility they have with regards to the program is greater than those in other program roles. Those in a supervisory or executive position are likely to be responsible for tracking of a number of programs; therefore the program evaluated and included in this study represents only a portion of their work responsibility and the time dedicated to the program managed accordingly. Program sponsors have a similar relation to the program. They sponsor a range of programs of which the program in this study is only one. Alternatively, those responsible for program delivery and program clients or participants have a clear stake in

the program, although their level of responsibility is not as high as those responsible for managing the program. For this reason, a decision was made to weight the role of Program Manager/ Co-ordinator by '2'. To illustrate the calculation of stakeholder stake in the program, some examples have been included below.

Items 6, 9 and 10, adapted from the work of Cummings (1997), focus on the stakeholder's perception of the need for the program, the efficacy of program delivery and the program philosophy. Items 7 and 8 were developed for this study to gather information regarding the stakeholders' perception of intended program outcomes and current manner of program service delivery.

Items 11, 12 and 13 focus on the stakeholder's sense of personal efficacy in relation to the program's development, delivery and outcomes. There is substantial literature available focusing on the links between teachers' perceived teaching efficacy, general teaching efficacy and the merits this perception brings to the work they do and the way they undertake it (Bandura, 1977; Guskey, 1988). This study has not attempted to determine in detail the stakeholder's sense of personal efficacy with regards to the program. The three items have been included to gain some insight into the stakeholder's sense of efficacy with regards to the program, based on the logic that a higher sense of efficacy with regards to the program leads to an increased sense of commitment to the program. Furthermore, Cousins and Walker (2000) found teachers' sense of personal efficacy to be a variable predictive of attitudes towards systematic inquiry, supporting the findings of previous studies indicating that high efficacy teachers are more receptive to new innovative, challenging, change-oriented, collaborative and academically-oriented practices and teaching techniques (Guskey, 1988; Smylie, 1988). For the purposes of this study, the link between perceived efficacy with relation to the program and use of evaluation information, will be explored.

Finally, items 14 and 15 were included to force stakeholders to commit to a choice of either support or opposition to the program. These items were specifically included by Cummings to force stakeholders to state a view either

supporting or opposing the program, and is underpinned by the work of McGuire (1969) and Kiesler (1971) indicating the more public a person's view is, the more likely they are to defend that view. Item 14 focusing on stakeholder support for the program requires respondents to choose between two options regarding their support or opposition to the program. Item 15 then asks them rate on the five-point scale, the strength of that choice as preponderance. In the item analysis, Item 15 has been modified to include only positive ratings regarding the stakeholder's support of the program. All negative ratings have been scored as '0'. Only Item 15 has been included in the scale item analysis.

#### PERCEIVED CHARACTERISTICS OF THE STAKEHOLDER COMMITMENT TO THE PROGRAM: SCALE ITEM ANALYSIS

Data analysis of the *Perceived Characteristics of the Stakeholder Commitment to the Program Scale* began with the inclusion of the 13 items listed in Table 5.6 below to determine stakeholder commitment to the program. The initial Cronbach's Alpha calculated was  $\alpha = 0.82$ , but with the deletion of Item 9 in the reliability analysis, the scale Alpha was increased to  $\alpha = 0.84$ . Inter-item correlations for items range from -0.14 to 0.71. Item 5, Item 7 and Item 8 have an  $r_{ii'} \geq 0.3$  with three items. Although these three items have low inter-item correlations values, the deletion of all three items from the scale would have only resulted in an increase of 0.01 in the Cronbach's Alpha value. Therefore, they have been retained as scale items for the value they add to the range of the scale. The other items have an  $r_{ii'} \geq 0.3$  with five items or more. The final scale of twelve items has a mean of 43.92 and a standard deviation of s.d.= 8.75. The results of the reliability analysis are shown in Table 5.6 below.

#### PERCEIVED CHARACTERISTICS OF THE STAKEHOLDER COMMITMENT TO THE EVALUATION STUDY SCALE

In a study focusing on the influence of stakeholder involvement in evaluation on the use of evaluation information, Cummings (1997) found stakeholder commitment to the evaluation study to have influence on use of the evaluation

information. Other authors concur (Alkin et al., 1979; Dickey, 1980; Hudson-Mabbs, 1993; Patton et al., 1977; Vlahov, 1989).

Nine items have been included in this section to gather information from stakeholders regarding their perspective of the evaluation study. The first item, Item 16, is an open response item focusing on stakeholders' views regarding the main purpose of the evaluation study. This item was adapted from Cummings (1997) work. Five items (items 17, 18, 19, 20, 21/22) have been included to form the Perceived Characteristics of the Stakeholder Commitment to the Evaluation Study Scale. The reliability analysis of the scale is summarised in the following section. Four items of the Perceived Characteristics of the Stakeholder Commitment to the Evaluation Study Scale of this study (Item 17, Item 18, Item 19 and Item 20) focus on the stakeholder's perception of the utility of the evaluation study, or the extent to which they expect the evaluation is suitable or adaptable to an end are based on items included in the works of Hudson-Mabbs (1993) and Cummings (1997). These items request responses from stakeholders on a five-point scale. Two items (Item 17- stakeholder perception of the need for the evaluation study and Item 21/Item 22- stakeholder support of the evaluation study) are based on items included in the work of Cummings (1997) to assess stakeholder commitment to the evaluation study as an indicator of evaluation information use.

As is the case with Item 14, Item 21 focusing on stakeholder support for the evaluation, requires respondents to choose between two options regarding their support or opposition of the evaluation. Item 22 then asks them to rate on the five-point scale the strength of that choice as preponderance. In the item analysis, Item 22 has been modified to include only positive ratings regarding the stakeholder's support of the program. All negative ratings have been scored as '0'. Only Item 22 has been included in the scale item analysis.



**Table 5.6: Perceived Characteristics of the Stakeholder Commitment to the Program. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{it} \geq 0.3^1$	$r_{it}^2$
<b>Program Involvement</b>			
ITEM 2	How have you been involved with the program?	6	0.48
ITEM 3	Were you involved with the early development of the program?	9	0.67
ITEM 4	Were you involved with the initial delivery of the program?	9	0.73
ITEM 5	Are you currently involved in the delivery of the program?	3	0.36
<b>Commitment to Program</b>			
ITEM 6	Do you feel there is a need for the program?	8	0.56
ITEM 7	Do you feel the intended outcomes of the program are worthwhile?	3	0.31
ITEM 8	Do you agree with the manner in which the program services/practices are delivered?	4	0.38
ITEM 9	Do you feel there are better ways of delivering this program? <b>Item deleted due to low <math>r_{it}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.01</math>). When item deleted <math>\alpha = 0.84</math>.</b>	1 <sup>3</sup>	
ITEM 10	Do you feel the program is based on a sound philosophy?	5	0.49
ITEM 15	How strongly do you support/ oppose the program?	8	0.55
<b>Personal Influence on Program</b>			
ITEM 11	Do you feel you have personally influenced the <b>development</b> of this program?	7	0.58
ITEM 12	Do you feel you personally influence the <b>delivery</b> of this program?	5	0.67
ITEM 13	Do you feel you personally influence the <b>outcomes</b> of this program?	7	0.56
Note:	<sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{it} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents the number of other scale items the item correlates with ( $r_{it} \geq 0.3$ ) in first run of scale with all items included.		

The final items of the section, items 23 and 24, focus on stakeholders' perception of whether they believe the evaluation study findings will be supportive or critical of the program. Item 23 asks for a dichotomous response (Support or Critical), while Item 24 is open response format, focusing on how they expect the findings to be supportive or critical of the program.

#### PERCEIVED CHARACTERISTICS OF THE STAKEHOLDER COMMITMENT TO THE EVALUATION STUDY: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the five items listed in the Table 5.7 to determine stakeholder commitment to the evaluation study. The alpha value of the final *Perceived Characteristics of the Stakeholder Commitment to the Evaluation Study Scale* is  $\alpha = 0.83$ . A review of inter-item and item-total correlations for the items indicate the basis for the inclusion of all five in the scale to assess stakeholder commitment to the evaluation study is sound inclusions. Inter-item correlations for items range from 0.21 to 0.71. Only Item 20 correlates ( $r_{ii'} \geq 0.3$ ) with two items. All other items correlate ( $r_{ii'} \geq 0.3$ ) with three items or more. The item-total values (corrected) for all items is good ( $r_{it} \geq 0.3$ ). The final scale of five items has a mean of 21.75 and a standard deviation of s.d.= 3.35. The results of the reliability analysis are presented below in Table 5.7.

#### PERCEIVED STAKEHOLDER EXPECTATION OF INVOLVEMENT IN THE EVALUATION STUDY SCALE

A positive relationship between stakeholder commitment to the evaluation, defined as involvement, and evaluation information use, has been identified by numerous authors (e.g. Brett, Hill-Mead and Wu, 2000; Cousins and Earl 1992, 1995; Cousins and Leithwood, 1986; Dawson and D'Amico, 1985; Cummings, 1993; Greene, 1988; Hudson-Mabbs, 1993; Patton, 1997; Preskill and Caracelli, 1997; Rich, 1979). Communication between those responsible

for the evaluation and the evaluation stakeholders as a particular tool of involvement has been, and continues to be, an area of interest (for example, Cummings, 1997; Greene, 1988). A number of reasons for the strong link between stakeholder involvement in the evaluation and use of its information have been identified. It has been suggested that the involvement of stakeholders in the evaluation process increases the likelihood the information needs of the stakeholders will be met, making the evaluation information more relevant to them (Cummings, 1997; Patton, 1997).

**Table 5.7: Perceived Characteristics of the Stakeholder Commitment to the Evaluation Study. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Evaluation Study Need</b>			
ITEM 17	Do you feel there is a need for this evaluation study?	4	0.74
ITEM 22	How strongly do you support this study? <sup>3</sup>	4	0.82
<b>Usefulness of Evaluation Study Information</b>			
ITEM 18	Do you expect this study will provide information useful to the program?	3	0.67
ITEM 19	Do you expect this study will provide information useful to _____(the organisation responsible for the program)?	3	0.52
ITEM 20	Is the program at a stage in its life where an evaluation is appropriate?	2	0.49
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Item reworded for presentation in table.			

It has also been suggested that the involvement of stakeholders in the evaluation process prepares them for accepting the information by making them aware of the program context and the evaluation processes. This involvement, it is proposed, increases stakeholder's: understanding of the evaluation findings; ownership of the evaluation information; sense of personal responsibility for advocating the evaluation information; and likelihood of accepting the information as valid and credible (e.g. Cummings, 1997; Greene, 1988; Preskill and Torres, 1999a). It has also been proposed that the involvement of stakeholders in the research/evaluation process helps to create interpersonal networks necessary for the sharing of information, meaningful discussion and reflection (Cousins and Earl, 1995; Louis and Simsek, 1991). Furthermore, strategies to increase stakeholder involvement in research and evaluation processes likely to increase stakeholder use of the study information, have also been linked to the development of organisational learning (Cousins 1996, Leithwood and Louis 1999).

The items included focusing on stakeholder's anticipation regarding involvement in all stages of the evaluation study, have been adapted from the work of Cumming (1997). Cummings included a version of these items in his *Post-Study Questions*. Here they have been adapted to gather information pre-study to gain insight into stakeholders' expectations regarding involvement in the evaluation study. Similar items, with changes to wording tenses of the items, have also been included to gather similar information from the stakeholders of the evaluations included in this study, following the dissemination of the relevant evaluation's final report (Stakeholder Interview Schedule II) with a view to comparing and contrasting the expectations of stakeholders regarding involvement in the evaluation study and their perception of the extent to which they were actually involved. Furthermore, it is expected that the more stakeholders expect to be involved in the evaluation, the more likely they will be involved.

The nine items have been included in the interview schedule to gain insight into stakeholder's expectations regarding their involvement in the evaluation

study. These nine items were intended to form the *Perceived Stakeholder Expectation of Involvement in the Evaluation Study Scale*.

Items included in the *Perceived Characteristics of the Stakeholder Involvement in the Evaluation Process Scale* focus on involvement in; planning (Item 25), conduct (Item 27), communication (Item 30, Item 32), elaboration of the program logic (Item 33), dissemination of the findings (Item 34), and development of the recommendations (Item 35). In addition, two items have been included to assess the extent to which the stakeholder's involvement in the evaluation study will be expected (Item 26, Item 28). The influential role the organisation has on the evaluation process and the use of the information, has been detailed by numerous authors (e.g. Alkin et al., 1979; Cousins and Earl, 1995; Patton et al., 1977). The two items included here have been included to gain insight into the influence the expectations of the organisation have on stakeholders' anticipation regarding their involvement in the evaluation. All items are on a five point scale.

Two items, Item 29 focusing on evaluator reporting in the evaluation process and Item 31, focusing on the stakeholder's desire to be interviewed in the evaluation, have not been included in this scale, as the nature of the items is such that the response format to these items post-evaluation study required dichotomous responses (yes/no), rather than responses on a five point scale. Therefore, they are not appropriate for a pre/post scale comparison. The data collected in response to these items is reported later in this paper to add insight to other qualitative and quantitative information gathered in this study. The analysis of the scale is summarised in the following section.

#### PERCEIVED STAKEHOLDER EXPECTATION OF INVOLVEMENT IN THE EVALUATION STUDY SCALE: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the nine items listed in Table 5.8 below to determine stakeholder expectations regarding their involvement in the evaluation study. Cronbach's Alpha was first calculated as  $\alpha = 0.84$ . In the reliability analysis, one item (Item 34) was deleted from the scale. Although

the deletion of Item 34 resulted in only a minor increase of the scale alpha, the item was deleted on the basis of generally low inter-item correlation values ( $r_{ij}$  = 0.3 with one item) and a low item-total correlation value ( $r_{it}$  = 0.06).

The alpha value of the final *Perceived Stakeholder Expectation of Evaluation Involvement Scale* is  $\alpha$  = 0.86. Final inter-item correlations for items ranged from 0.15 to 0.70. All final scale items have a  $r_{ij} \geq 0.3$  with four other scale items or more. The final scale of eight items has a mean of 24.72 and a standard deviation of s.d. = 7.54. A summary of a reliability analysis of the *Perceived Stakeholder Expectation of Involvement in the Evaluation Study* is presented in Table 5.8.

#### **PERCEIVED STAKEHOLDER CHARACTERISTICS- EDUCATION AND TRAINING IN SOCIAL RESEARCH AND EVALUATION SCALE**

Studies indicate that previous education and experience of research positively influences stakeholder opinions of social research, their perceptions regarding their understanding of it and ability to undertake it (Cousins and Walker, 2000; Green and Kvidahl, 1990). Furthermore, previous participation in research was found by Cousins and Walker (2000) to be a significant predictor of attitude towards research. Alternatively, Patton (1997) writes of concerns that stakeholders with a limited research background may be intimidated by the conceptualisations and terminology adopted in a theory-based evaluation, thereby decreasing the likelihood that the evaluation information will be used. It is expected previous education and experience with social research and evaluation will have a positive influence on stakeholder openness to using evaluation information.

Cousins and Walker (2000) found that when educators had undertaken prior research coursework they had more favourable attitudes towards research. The present study takes this finding a step further by focusing on the link between stakeholder education, training and experience with social research and evaluation and the use of the evaluation information. Items 36 through 41 focus on stakeholders' perception of their educational background (Item 36)

and training or experience (Item 38). Items 36 and 38 require a response on a five point scale. In addition, four open response items have been included to gather further information regarding a summary of education background (Item 37), a summary of experience or training (Item 39), the number of research projects they have been involved with (Item 40) and the manner in which they were involved (Item 41).

**Table 5.8: Perceived Characteristics of the Stakeholder Involvement in the Evaluation Process Scale. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Expected Involvement in the Evaluation Process</b>			
ITEM 26	Will you be <u>expected</u> to be involved in planning the evaluation study (e.g. by your own organisation, the organisation responsible for the program)?	6	0.69
ITEM 28	Will your involvement in the conduct of the evaluation study be <u>expected</u> ?	4	0.51
<b>Desired Involvement in the Evaluation Process</b>			
ITEM 25	Would you <u>like</u> to be involved in <u>planning</u> the evaluation study?	6	0.70
ITEM 27	Would you <u>like</u> to be involved in the <u>conduct</u> of the evaluation study?	5	0.70
ITEM 33	Would you like to be involved in any work to understand or explain the logic, rationale or theory underpinning the program?	6	0.62
ITEM 30	Would you <u>like</u> to be in <u>regular communication</u> with the evaluation team?	7	0.77
ITEM 32	Would you <u>like</u> to be able to <u>feedback</u> to the evaluation team on issues and concerns regarding the evaluation?	4	0.47
ITEM 34	Would you <u>like</u> to receive the evaluation findings as they emerge?		
<b>Item deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.06</math>). When item deleted <math>\alpha = 0.86</math>.</b>		1 <sup>3</sup>	
ITEM 35	Would you <u>like</u> to be involved in <u>developing the recommendations</u> from the study?	4	0.41
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ) in first run of scale with all items included.			

Items 45 to 54 are similar to the items of the previous sub-section except they target education, training and experience with evaluation studies rather than social research. The exception is the addition of one item to gain insight into the stakeholders' perception of evaluation (Item 54). Two five-point scale items have been included which focus on stakeholders' perception of the education (Item 45) and experience or training (Item 47) regarding evaluation. Four open-response items have been included to gather further information regarding a summary of evaluation-relevant education background (Item 46), a summary of experience or training regarding evaluation (Item 48), the number of evaluation projects they have been involved with (Item 49) and the manner in which they were involved (Item 50). Again, although the responses from these items have not been included in this scale, the information will be reported later in this study to provide further insight regarding the education, experience and training of stakeholders with evaluation.

Four items have been included in the interview schedule to gain insight into stakeholders' background in terms of education and training in social research and evaluation. These four items (Items 36, 38, 45 and 47) were intended to form the *Perceived Stakeholder Characteristics- Education and Training in Social Research and Evaluation Scale*. The reliability analysis of the scale is summarised in the following section.

#### PERCEIVED STAKEHOLDER CHARACTERISTICS- EDUCATION AND TRAINING IN SOCIAL RESEARCH AND EVALUATION: SCALE ITEM ANALYSIS.

Data analysis began with the inclusion of the four items listed in Table 5.9 below to determine stakeholder characteristics of education and training in social research and evaluation. Cronbach's Alpha of all four items was  $\alpha = 0.85$ . A review of inter-item and item-total correlations ( $r_{it} \geq 0.03$ ) indicate all four items are sound inclusions in the scale to assess the stakeholder characteristic of education and training, or experience with social research and evaluation. Inter-item correlations for items ranged from 0.49 to 0.77. All items correlate with all other items ( $r_{ij} \geq 0.3$ ). The final scale of four items has



a mean of 11.89 and a standard deviation of s.d.= 3.9. The results of the reliability analysis are presented below in Table 5.9.

**Table 5.9: Perceived Stakeholder Characteristics: Education and Training in Social Research and Evaluation Scale**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Education, Training and Experience with Social Research</b>			
ITEM 36	To what extent has your educational background included some exposure to social research?	3	0.71
ITEM 38	To what extent has your training or experience included some exposure to social research?	3	0.61
<b>Education, Training and Experience with Evaluation</b>			
ITEM 45	To what extent has your educational background included any exposure to evaluation?	3	0.75
ITEM 47	To what extent has your training or experience included some exposure to evaluation?	3	0.71
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.			

#### PERCEIVED OPPORTUNITY OF STAKEHOLDERS TO USE INFORMATION SCALE

The opportunity the stakeholder has to access the research and evaluation information, to reflect upon the information and opportunity to apply the information, represent the variable Opportunity to Use Information in this study. Items 42 to 44 of the section focus on the extent to which respondents are able to read social research reports (Item 42), consider the implications for their work (Item 43), and apply the information to their work (Item 44). These items were developed to gather information regarding stakeholders' opportunity for social research information use. Items 51, 52 and 53 have the same focus with regards to use of evaluation information. One item (Item 54) has also been included to gain insight into the stakeholders' perception of

evaluation. This item does not fit with the scale, though the responses have been used to add further insight to study findings.

Six items have been included in the interview schedule to gain insight into stakeholders' perceptions regarding opportunities to use information. These six items (items 42, 43, 44, 51, 52, and 53) were intended to form the *Perceived Opportunity of Stakeholders to Use Information Scale*. Simply, for use of information to occur, both conceptual and operational opportunities to attend to and apply the information must be available to the individuals; only then is larger scale use of information likely to occur (Kim, 1993; Preskill and Torres, 1999a). These items are a reflection of the stakeholders' perception of the organisation culture within which they work. The items ask the extent to which respondents are "able" to read reports, consider the implications of study information and apply study information to their work, again recognising the powerful influence of the organisation culture and norms on the access and use of research and evaluation information (e.g. Cousins and Walker, 2000; Cousins and Earl, 1995; Preskill and Torres, 1999b). The reliability analysis of the scale is summarised in the following section.

#### PERCEIVED OPPORTUNITY OF STAKEHOLDERS TO USE INFORMATION: SCALE ITEM ANALYSIS.

Data analysis began with the inclusion of the six items listed in Table 5.10 below to determine the opportunity of stakeholders to use social research and evaluation information. Items 42 and 51 focus on the extent to which the stakeholders are able to read social research and evaluation reports, respectively. Items 43 and 52 focus on the extent to which stakeholders are able to consider the implications of social research and evaluation information for their work. Items 44 and 53 are concerned with the extent to which they are able to apply the social research and evaluation information to their work.

The alpha value of the final *Perceived Opportunity of Stakeholders to Use Information Scale* is  $\alpha = 0.80$ . A review of inter-item and inter-total correlations ( $r_{it} \geq 0.3$ ) indicated all six are sound inclusions in the scale to

assess stakeholder opportunity to use information. Inter-item correlations for items ranged from 0.01 to 0.70. Only Item 53 has an  $r_{ii'} \geq 0.3$  with three items. All other items have an  $r_{ii'} \geq 0.3$  with four items or more. The final scale of six items has a mean of 21.86 and a standard deviation of s.d.= 4.176. The results of the reliability analysis are presented below in Table 5.10.

**Table 5.10: Perceived Opportunity of Stakeholders to Use Information Scale**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Opportunity to Attend to Information</b>			
ITEM 42	To what extent have you been able to read social research reports or articles?	3	0.45
ITEM 51	To what extent have you been able to read <i>evaluation</i> reports or articles?	4	0.57
<b>Opportunity to Consider Implications of Information for Work</b>			
ITEM 43	To what extent are you able to consider the implications of research study information for your work?	4	0.61
ITEM 52	To what extent have you been able to consider the implications of <i>evaluation</i> information for your work?	4	0.66
<b>Opportunity to Apply Information to Work</b>			
ITEM 44	To what extent are you able to apply research study information to your work?	4	0.55
ITEM 53	To what extent have you been able to apply <i>evaluation</i> information	3	0.50
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.			

## PERCEIVED STAKEHOLDER COMMITMENT TO THE ORGANISATION

The logic linking stakeholder commitment to the organisation with use of the evaluation information, is an extension of the work of Cumming (1993) in which he included items focusing on stakeholder commitment to the program and stakeholder commitment to the evaluation study. Both of these variables have been included as indicators of evaluation information use in this study.

The influence of the organisation culture on the use of the evaluation information is a third independent variable of this work. Therefore, the inclusion of a scale focusing on stakeholder commitment to it is necessary.

Adopted from the work of Preskill and Torres (1999b), the six items to determine stakeholder commitment to the organisation are concerned with stakeholders' perception of: the strategic direction of the organisation (Item 55); the need for organisation services (Item 56); the efficacy of organisation service delivery (Item 57); the culture of the organisation (Item 58); the organisation's capacity to change (Item 59); and, their personal efficacy in relation to the organisation (Item 60). All items require a response on a five point scale. These six items were intended to form the *Perceived Stakeholder Commitment to the Organisation Scale*. The scale is composed of two sub-scales: The *Perception of Organisation Direction and Need Sub-scale*; and *The Perception of Organisation Culture, Capacity to Change, and Personal Efficacy Sub-scale*. The reliability analysis of the scale is summarised in the following section.

#### PERCEIVED STAKEHOLDER COMMITMENT TO THE ORGANISATION: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the six items listed in the Table 5.11 below to determine stakeholder commitment to their organisation. The initial Cronbach's Alpha ( $\alpha$ ) calculated was  $\alpha = 0.61$ . A review of inter-item and inter-total correlations of all the items encouraged the deletion of Item 57 due to low inter-item correlation values ( $r_{ii} \geq 0.3$  with no items). The deletion of Item 57 resulted in Cronbach's Alpha increasing to 0.63. However, the inter-item and inter-total correlations of the remaining five items indicated the items were measuring two sub-dimensions. Analysis of the items as two separate sub-scales supported this decision. Therefore, Item 55 and Item 56 comprise Sub-scale 1 (*Perception of Organisation Direction and Need*) while Item 58, Item 59, and Item 60 together form Sub-scale 2 (*Perception of Organisation Culture, Capacity to Change, and Personal Efficacy*). Sub-scale 1 and Sub-

scale 2 together represent the final *Perceived Stakeholder Commitment to the Organisation* scale.

The Cronbach's Alpha of Sub-scale 1 (*Perception of Organisation Direction and Need*) is  $\alpha = 0.69$ . The inter-item correlation ( $r_{ii'}$ ) for the two scale items (Item 55 and Item 56) is 0.54. The items correlate with each other ( $r_{ii'} \geq 0.3$ ). The final scale of five items has a mean of 8.83 and a standard deviation of s.d.= 1.21.

The Cronbach's Alpha of Sub-scale 2 (*Perception of Organisation Culture, Capacity to Change, and Personal Efficacy*) is  $\alpha = 0.73$ . Inter-item correlations ( $r_{ii'}$ ) for items ranged from 0.37 to 0.61. The three sub-scale items (Item 58, Item 59 and Item 60) correlate with each other ( $r_{ii'} \geq 0.3$ ). The item-total correlations for all three items are good ( $r_{it} \geq 0.3$ ). The final scale has a mean of 10.92 and a standard deviation of s.d.= 2.1. A summary of the two sub-scales is presented in Table 5.11.

## THE FINAL SECTION

The final section of the Stakeholder I Interview Schedule includes two five-point scale response items focusing on the formalised evaluation activities incorporated into the delivery of the program (Item 63) and the information use routines in decisions about the program. These two items alone do not represent a scale. However, the data of these items is considered later in this study.

In addition, four open response items have been included in this section to gather information regarding when the last formal evaluation of the program was undertaken (Item 61), if a formal evaluation of the program has been undertaken recently (Item 62) and the sort of information the respondent refers to in making decisions regarding the program (Item 65). The information collected with regards to these items is considered later in this study.

**Table 5.11: Perceived Stakeholder Commitment to the Organisation Scale <sup>1</sup>**

Scale Items Grouped by Area of Focus		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Sub-scale 1: Perception of Organisation Direction and Need</b>			
ITEM 55	Do you agree with the main strategic direction of the organisation?	1	0.54
ITEM 56	Do you feel there is a need for the services/ programs of this organisation?	1	0.54
<b>Sub-scale 2: Perception of Organisation Culture, Capacity to Change, and Personal Efficacy</b>			
ITEM 58	Do you feel the organisation has a positive culture?	2	0.65
ITEM 59	Do you believe the organisation has the capacity to change if more effective methods of program/service delivery are identified?	2	0.59
ITEM 60	Do you feel your role in relation to the organisation is effective?	2	0.46
<b>Item not included in either sub-scale.</b>			
ITEM 57	Do you feel there are better ways of delivering organisation services?		
<b>Note:</b>	<sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.		

<sup>1</sup> The Items of this scale are grouped into two separate sub-scales, not one scale represented by areas of Item focus as in the previous tables.

## SUMMARY

The six sections of the Stakeholder I Interview Schedule include a mix of items of both open and scaled response format. A number of items have been adapted from the work of Cumming (1997) and others, as identified. Responses to the open response items and scaled responses items not included in a scale, is used to add further insight to the quantitative findings based on the scales later in this work.

The six scales of the interview schedule detailed above represent the perceived stakeholder characteristics included in this study which are expected to have

influence on the use of the information of the evaluations included in this study. Interview data was used for analysis. With regards to each scale, a few items with weak inter-item correlation ( $r_{ii'} \leq 0.03$ ) have been deleted. All study scales have been verified (Cronbach's Alpha  $\geq 0.7$ ) with the exception of the two sub-scales of the *Perceived Stakeholder Commitment to the Organisation Scale*. Cronbach's Alpha of both sub-scales is  $\alpha \geq 0.6$ . The analysis of the scales helped confirm what was anticipated in terms of scale and sub-scale content and focus.

#### EVALUATOR INTERVIEW SCHEDULE I

As we write, managers and administrators have become the largest group of employees in Britain. There are 3,921,000 of them, 15.8% of the labour force, all charged with making judgments to better their organisations. Here lies the great promise of evaluation: it purports to offer the universal means with which to measure 'worth' and 'value'. Evaluation, in short, confers the power to justify decisions" (Pawson and Tilley 1997; p xii).

Evaluators are responsible for undertaking evaluations with the potential for powerful influence. Weiss's (in Alkin, 1990) belief that the interest, ideology and background of an evaluator impact on the use of evaluation information, is supported by numerous authors (Alkin and Dailak, 1985; Cousins and Leithwood, 1986; Greene, 1988; Hammond, 1983; Mathison, 1994; Seigel and Tuckel, 1985). In fact, Hudson-Mabbs (1993) found the credibility of the evaluator and the evaluation processes were particularly important where potential dissension exists between interested parties.

Evaluators are, for the purposes of this investigation, considered to be anyone responsible for managing or undertaking either all of the evaluation process or a specific task(s) within it. Although the primary evaluator or evaluation manager has considerable opportunity to influence the evaluation, other members of the evaluation team also have an influence through the

undertaking of their duties. The decision to interview all layers of the evaluation team was made to account for the potential of the evaluation team to influence the evaluation individually, as well as collectively.

Two interview schedules were developed to gather information from the evaluators of each case: the Evaluator I Interview Schedule and the Evaluator II Interview Schedule. Both are detailed in the sections below.

The purpose of the Evaluator Characteristics Interview Schedule I is to gather information from those responsible for undertaking the evaluations included as cases in this study. The five sections of the Evaluator Characteristics Interview Schedule I of this study focus on: the role the evaluator has had in the undertaking of the evaluation included in this study and how their involvement came about; their evaluation and research background in terms of training, education, work experience and commitment to the field; the evaluators' perception of the importance they place on the various roles an evaluator plays in the course of an evaluation; and their thinking regarding various aspects of evaluator practice.

The purpose of this section is to describe the items of the interview schedule and provide details of the reliability analysis of the quantitative scale data collected in the administration of the Evaluator I Interview Schedule. The items included in the interview schedule sections are a mix of open response and scaled response format. The scales and sub-scales of the interview schedule and the items contributing to them will be further discussed in the sections below and the details given of the reliability analysis of the four scales (*Evaluator Perception of Evaluator Practice* Sub-scales 1, 2 and 3, represented in Table 5.12, and *Perceived Characteristics of Evaluator Regarding Stakeholder Involvement in the Generic Evaluation Process Scale*, Table 5.13). In addition, two instruments developed for this study to gain insight into the evaluators' perceptions of evaluator practice have been included in this interview schedule.



## EVALUATION INVOLVEMENT INSTRUMENT

Numerous authors consider the influence of the characteristics of the evaluator or evaluation team on the evaluation process and outcomes (e.g. Braskamp et al., 1982; Conley-Tyler, 2005; Cummings et al., 1988; Greene, 1988; House and Howe, 1998; Lake, 2005; Mathison, 1994; Owen, 2006; Scriven, 1991; Weiss, 1972). The first two items of this interview schedule are open response items intended to: gather information regarding the role of the respondent with regards to the evaluation; identify the location of the respondent to the program or organisation (e.g. external or internal); and gain some insight into the process (if any) by which they became involved in the evaluation (e.g. tender process, previous work). Item 1 focuses on the role the respondent plays in the evaluation team. Item 2 deals with how they became involved in the evaluation of the program (Item 2.a.) and, if they were chosen as the result of a selection process, why they feel they were selected (Item 2.b.). These items are expected to yield useful insight into the history between the evaluation team, the program evaluated, and the organisation and the factors the evaluators of the evaluation studies included in this study consider important in instigating their involvement in the evaluation.

A scale to represent the nature and degree of the interviewee's involvement (Item 1) in the evaluation was developed for this study and is discussed below. The first section deals with the rated area of involvement and the characteristics of each. This is followed by a brief description of each interviewee by evaluation. Three tables and two figures summarising the evaluation involvement rating scale and the interviewee scores are included. The section concludes with a discussion about the issues considered in the application of the rating scale.

## RATING OF EVALUATION INVOLVEMENT

Four particular areas of involvement in the evaluation were considered. These areas are; the evaluation conceptualisation, the evaluation process, evaluation process management and the evaluation reporting process. Interviewee

responses were rated in the four areas and totalled to produce an overall rating of involvement in the evaluation study. Each of the four areas has four rating levels ('0', '1', '2', and '3'). The four rated areas and the rating characteristics are detailed more fully below. Table 5.12 summarises the rating characteristics.

#### EVALUATION CONCEPTUALISATION

Evaluation conceptualisation is considered to include involvement in the early meetings to negotiate and delineate the evaluation strategy and process. Meetings attended were for the purpose of; delineating the evaluation strategy, undertaking program theory explication, determining information focus, and instrument development. The level of active participation in the initial evaluation conceptualisation meeting determines the primary difference between the ratings of '0', '1', '2', and '3'. If they did not attend the meetings a rating of '0' is appropriate. If the meetings were attended, but the interviewee tended to adopt a more passive role as opposed to a more active and influential role, they were given a rating of '1'. Interviewees who attended the meetings and were actively influential in the negotiation process were given a rating of '2'. Those who attended the meetings and tended to actively guide the negotiation of the evaluation strategy, program theory elaboration and/or instrument development, were given a rating of '3'. Confident evaluators with a great deal of evaluation experience fell into this category.

The scores of only two categories of involvement have been weighted: the evaluation conceptualisation and the reporting process. In both cases the score has been multiplied by 2. The logic underpinning the weighting is based on the assumption that the stages of evaluation conceptualisation and reporting are particularly important in defining the purpose, strategy and interpretation of the evaluation study, and therefore those involved at these times are particularly influential. Those involved in the undertaking of the evaluation certainly have the ability to influence the study, though the more practical tasks are heavily influenced by the evaluation direction determined before, and the interpretation of what occurs in, the evaluation and its findings.

**Table 5.12: Evaluator Role Ratings**

<b>Evaluator Role Scale Information</b>	
<i>Scale Focus</i>	<i>Definition of Focus</i>
<b>Evaluation Conceptualisation</b>	<p>0 Meetings, negotiation of evaluation strategy and process, theory explication</p> <p>1 No or minimal involvement in this area.</p> <p>2 Attended meetings - avenue for some influence on the evaluation process.</p> <p>3 Attended meetings to conceptualise evaluation strategy, process, instrument development, and program theory. Explication- significant influence on the evaluation process.</p> <p>4 Headed meetings and guided the negotiation of evaluation process and strategy conceptualisation, instrument development and program theory explication.</p>
<b>Evaluation Process</b>	<p>0 Data collection, data entry, data analysis</p> <p>1 No or minimal involvement in this area</p> <p>2 Undertook tasks related to data collection, data entry or data analysis</p> <p>3 Responsible for data collection, development of data query tools</p> <p>4 General overall responsibility for evaluation process</p>
<b>Evaluation Process Management</b>	<p>0 Evaluation administration and data management</p> <p>1 No or minimal involvement in this area</p> <p>2 Responsible for data collection and/or management</p> <p>3 Responsible for management of evaluation process-internal</p> <p>4 Responsible for management of evaluation process-external</p>
<b>Reporting Process</b>	<p>0 Interpreting, Drafting, Presenting</p> <p>1 No or minimal involvement in this area</p> <p>2 Undertook interpretation of the data and/or drafting the final report</p> <p>3 Audience for the evaluation findings have an avenue to influence final report draft</p> <p>4 Responsible for presentation of evaluation findings to stakeholders</p>

## EVALUATION PROCESS

The ratings developed for the level of involvement in the evaluation process refer to the practical tasks of the evaluation including data collection, data entry and data analysis. They are directly linked to the level of responsibility the interviewee had in the evaluation process. No or minimal involvement in the evaluation process was given a rating of '0'. Responsibility for undertaking discrete tasks such as data collection, data entry or analysis was given a rating of '1'. In cases where the interviewee was responsible for significant data collection or data analysis a rating of '2' was given. A rating of '3' was allocated to those with overall responsibility for managing the evaluation process. Those contracted as the primary evaluators were also given this rating.

## EVALUATION PROCESS MANAGEMENT

Ratings for this area were developed with regard to the level and type of responsibility for the management of the evaluation process. Those with no or minimal responsibility in this area were rated a '0'. Those responsible for managing administration of practical evaluation undertakings, such as data collection or data analysis were given a rating of '1'. A rating of '2' was appropriate for those responsible for the management of the evaluation process from within the organisation (i.e. maintaining communication with those interested in the evaluation, answering queries from the external contracted evaluators, etc.). A '3' was allocated to evaluators responsible for overall management of the evaluation process and maintaining the integrity of the evaluation process as conceptualised.

## REPORTING PROCESS

The final category in which the interviewees were rated refers to involvement and responsibility for the evaluation reporting process. Those with no responsibility in this area were given a rating of '0'. Interviewees who undertook some initial interpretation of the evaluation findings and/or drafting

of the report were allocated a '1'. Interviewees who were recipients of the evaluation findings (i.e. internal evaluation stakeholders) and had no means to influence the final interpretation of the findings were appropriated a score of '2'. Finally, those with primary responsibility for the final interpretations of evaluation findings and for justifying them to evaluation stakeholders were given a rating of '3'.

As mentioned above, the Reporting Process scores have been multiplied by '2'. The decision to weight this category of use has been based upon the significant influence of those involved in this stage of the evaluation study in the final interpretation of the evaluation findings.

### **Brief Description of Evaluator Characteristics Interview Schedule I Interviewees by Evaluation**

Those responsible for undertaking the evaluations are briefly described below by case. Where possible, a little history of previous program or evaluation involvement has been included with a view to contextualise the perception of each person described. Fictitious names have been used to ensure the confidentiality of each person.

### **Pre-Apprenticeship Evaluation Evaluator Involvement**

As the manager of the research and evaluation section of the Department of Education and Training, Christian Kick was responsible for internal management of the evaluation. Some of his responsibilities included a say in the engagement of contacting of outside consultants to undertake the evaluation. He also acted as an internal point of contact for the evaluators to provide program information and follow-up on requests. Christian played a role in instigating the evaluation of the Pre-Apprenticeship program. He was involved four years previously in a review of the Pre-Apprenticeship system that highlighted some gaps in program implementation. Adjustments of the program lead to the emergence of other issues to be considered. Christian's awareness of this program history and current issues caused him to play a role

in encouraging the evaluation of the Pre-Apprenticeship program. Midway through the study he was assigned to another position in the organisation and his duties regarding the evaluation were reassigned.

Zoë Roads (ZR) submitted the proposal for her research agency to undertake the Pre-Apprenticeship evaluation and was primary researcher contracted to undertake the evaluation. Although others at the agency worked on data analysis for the evaluation, she assumed primary responsibility for interviewing and other data collection, data analysis and reporting. She has previously been involved with research for the Department of Education and Training as part of her work at the research agency. Zoë wrote the entire final report of the evaluation and had full responsibility for presenting the findings to the stakeholders.

### **YOHFest Program Evaluation Evaluator Involvement**

A friend and colleague who was also a stakeholder of the program made Connor Fernandez (CF) aware of the need for an evaluator to undertake the evaluation of the YOHFest program. He submitted an evaluation proposal and was chosen to undertake the study. Connor was the primary evaluator throughout the study. He led all stakeholder meetings in the elaboration of the program logic, the negotiation of evaluation focus and development of instruments. He supervised all evaluation work and was responsible for the final report of the study. He had sole responsibility for taking the evaluation findings to the stakeholders.

Kim Apple (KA) was originally employed by the agency contracted to undertake the evaluation as an assistant to Connor Fernandez, the primary evaluator of the YOHFest program study. Therefore, she became involved in the evaluation by virtue of her position. She had a strong research background and was learning about evaluation. She was involved in all evaluation meetings to explicate the program logic and set up the evaluation. Other duties she undertook included data collection, data analysis and synthesis of the findings.

## **Making Consistent Judgments (MCJ) Evaluator Involvement**

Cole Right (CR) was the principal evaluator initially responsible for coordinating and overseeing the evaluation of the MCJ program. He chose midway through the study to hand the responsibility of the evaluation to Alex Hall (AH) due to his own demanding work schedule. However, Cole Right led the early meetings with the internal stakeholder team which was responsible for program management in the negotiation of the study focus, performance indicators to be used, the instrument development and evaluation strategy.

Alex Hall had some involvement in the initial conceptualisation of the evaluation. She has an evaluator role at the agency responsible for the evaluation contract and therefore is involved in most evaluations conducted by the agency. Midway through the study she assumed the role of primary evaluator for the final stages of the study. In her role she managed the study data analysis and final writing of the report. She had full responsibility for taking the study findings to the stakeholders.

Craig Baker (CB) was an employee of the agency contracted to undertake the evaluation. He was assigned to the MCJ study as part of his work for the agency. His considerable background in information management was put to use in undertaking data collection, data entry, setting up of the statistical query tools, data analysis (quantitative and qualitative), and in writing the report of the MCJ evaluation.

Gary Johnson, as manager of the directorate responsible for the MCJ program, has an interest in the program and its evaluation. He was involved in the selection of the external agency contracted to undertake the evaluation and in the discussions to negotiate the evaluation conceptualisation. The initial primary evaluator contracted to undertake the study has worked with Gary on previous evaluation studies.

Peta Patterson became involved in the MCJ evaluation by virtue of her position, as manager of a parallel program. She was a member of the internal

stakeholder team responsible for the management of the program and involved in the early discussions regarding the conceptualisation of the evaluation. Peta also acted as an internal point of contact for the contracted evaluation team, providing information and following up on requests.

Carla May (CM) became involved in the MCJ program evaluation primarily by virtue of her position though she also requested to be involved to further her evaluation experience. She assisted with the development of the tender to contract an outside evaluation group to conduct the evaluation. She also had input into evaluation conceptualisation as a member of the internal stakeholder team responsible for program management.

### **Discussion of the Rating Scale Application**

The ratings for involvement in each stage of the evaluation were, as a rule, easily determined from the information gathered in the first section of the interview schedule. The ratings given to each person with regards to each type of involvement is summarised in Table 5.13. In the determination of the rating, it was felt enough information had been gathered to enable a good understanding of where each person fell in each category. Generally, the highest rating of '3' was given to the primary evaluators contracted to undertake the evaluations in the *Evaluation Conceptualisation* and *Reporting Process* types of involvement. The exception is Christian Kick, the internal evaluation manager of the Pre-Apprenticeship program evaluation. In this case, Christian had undertaken a previous review of the program and as a result was quite knowledgeable of the program issues of concern to the Department of Education and Training. He also had a reasonable amount of experience with involvement in previous evaluations. By contrast, the external evaluator contracted to undertake the Pre-Apprenticeship program evaluation had never before undertaken an evaluation study, although she had considerable experience in the conduct of other types of research studies.

The primary evaluators contracted to undertake the YOHFest program and MCJ program evaluations were practiced evaluators. The experience of having



undertaken many evaluations previously, enabled them to assume a stronger role in the *Evaluation Conceptualisation* and *Reporting Process* stages of the evaluations with confidence. They stood out easily as the only candidates with a rating of ‘3’ in these types of use for both of these evaluations.

A rating of ‘2’, again in both of the *Evaluation Conceptualisation* and *Reporting Process* types of involvement, was generally given to internal program stakeholders or members of the evaluation team who had attended meetings for evaluation conceptualisation or reporting. Again, all interviewees given a two in these types of involvement stated they had influenced the meetings for these purposes in some way.

**Table 5.13: Involvement Rating Calculation and Final Scores**

<b>Ratings for each person by type of Involvement*</b>					
<b>Person (Program)</b>	<b>Evaluation Conceptualisation Rating (x 2 weighting)</b>	<b>Evaluation Process</b>	<b>Evaluation Process Management</b>	<b>Reporting Process Rating (x 2 weighting)</b>	<b>Final Involvement Rating</b>
<b>Making Consistent Judgments Program</b>					
<b>KA</b>	2 x 2= 4	3	1	1 x 2= 2	<b>10</b>
<b>CF</b>	3 x 2= 6	3	3	3 x 2= 6	<b>18</b>
<b>CB</b>	1 x 2= 2	2	1	1 x 2= 2	<b>7</b>
<b>CM</b>	1 x 2= 2	1	1	2 x 2= 4	<b>8</b>
<b>PP</b>	2 x 2= 4	1	2	2 x 2= 4	<b>11</b>
<b>GJ</b>	2 x 2= 4	0	2	2 x 2= 4	<b>10</b>
<b>YOHFest Program</b>					
<b>AH</b>	3 x 2= 6	3	3	3 x 2= 6	<b>18</b>
<b>CR</b>	3 x 2= 6	3	3	0 x 2= 0	<b>12</b>
<b>Pre-Apprenticeship Program</b>					
<b>ZR</b>	3 x 2= 6	3	3	3 x 2= 6	<b>18</b>
<b>CK</b>	3 x 2= 6	1	2	0 x 2= 0	<b>9</b>
* Higher ratings assume attainment of lower rating characteristics for that type of involvement.					

With regards to *Evaluation Conceptualisation*, two interviewees were given a '1'. In both cases they reported that they had not been in a position to heavily influence the conceptualisation process. Similarly, two interviewees were given a rating of '1' for the reporting process. Both were members of the evaluation team who had had considerable involvement in the evaluation process, though not in the negotiation or interpretation of the evaluation findings reported.

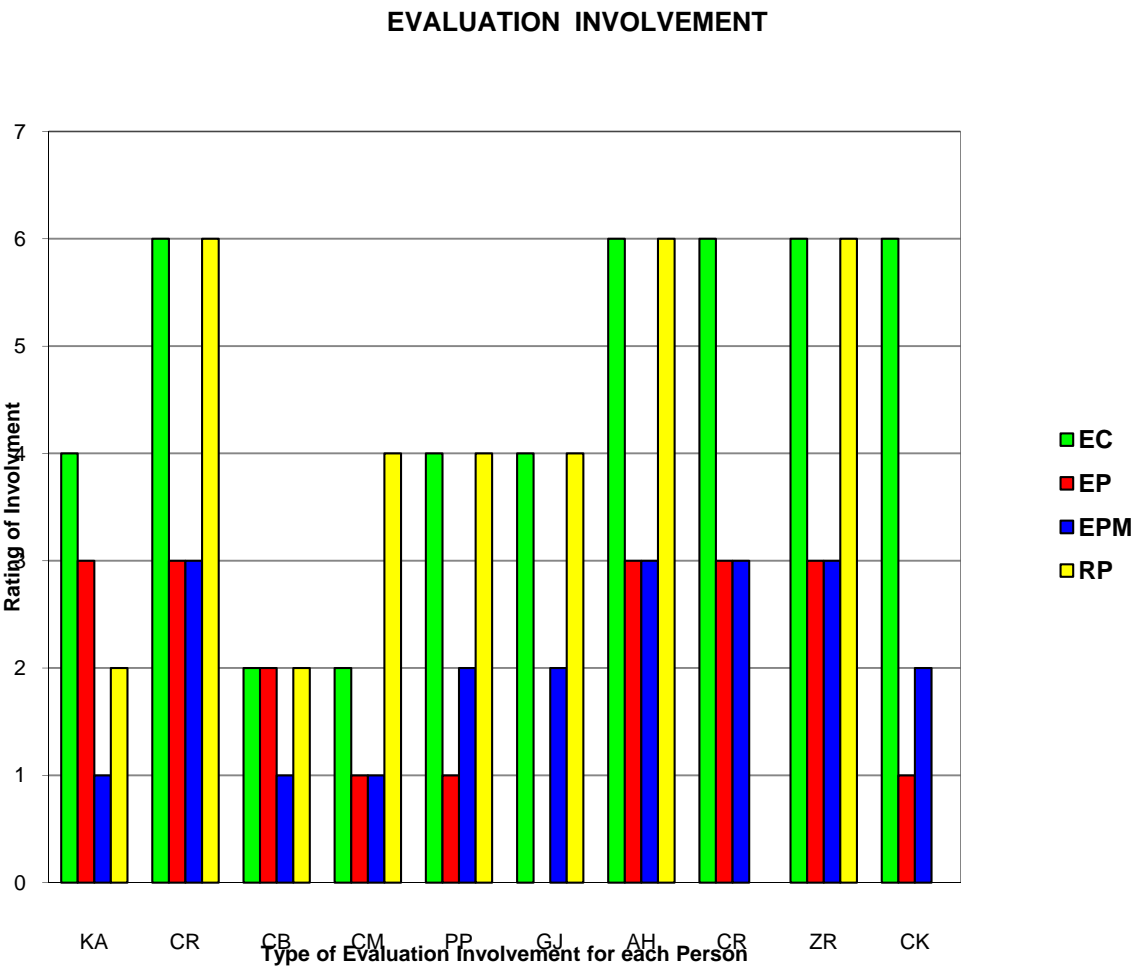
Ratings of '3' for involvement in the Evaluation Process were given to external evaluation team members for all three evaluations. The exception was Craig Baker of the MCJ program evaluation who was given a '1'. Although he was considerably involved in the evaluation process, he undertook tasks given to him due to his position at the agency contracted to undertake the evaluation. By contrast, Kim Apple worked very closely with the primary evaluator contracted to undertake the YOHFest program evaluation and was involved in all facets of the evaluation. Internal program stakeholders who acted as points of information for the evaluation were given a rating of '1'. One MCJ program stakeholder was given a rating of '0' as he had no involvement in the evaluation process.

Primary evaluators were given a '3' for Evaluation Process Management Involvement. In all cases it was their sole duty to maintain the integrity of the evaluation study design. A rating of '2' was given to internal program stakeholders with a particular vested interest in the evaluation. A '1' was given to members of the evaluation team primarily responsible for data collection and/or management. Finally, one internal stakeholder of the MCJ program was given a '1' due to her limited involvement with the management of the evaluation process.

The Evaluation Involvement Total Score by Person (Table 5.14) represents the total for each person resulting from weighting and adding the ratings. The three primary evaluators involved in the evaluations from beginning to end have the highest scores (18). The three lowest scores are the MCJ program evaluator team member who worked on components of the evaluation (7), a

MCJ program internal stakeholder with a less immediate influence on the program evaluation than others involved (8), and the internal evaluation manager of the Pre-Apprenticeship evaluation who was given a ‘0’ for the reporting process due to his deployment to another position in the department (9). The other interviewees tend to fall in the middle with a final rating of between 10 and 12. These four interviewees all had responsible involvement in the evaluation process, though were clearly not as influential as the primary evaluators. Table 5.15 includes a graph of interviewee Evaluation Involvement Characteristic ratings and totals, respectively.

**Table 5.14: Evaluation Involvement by Person**



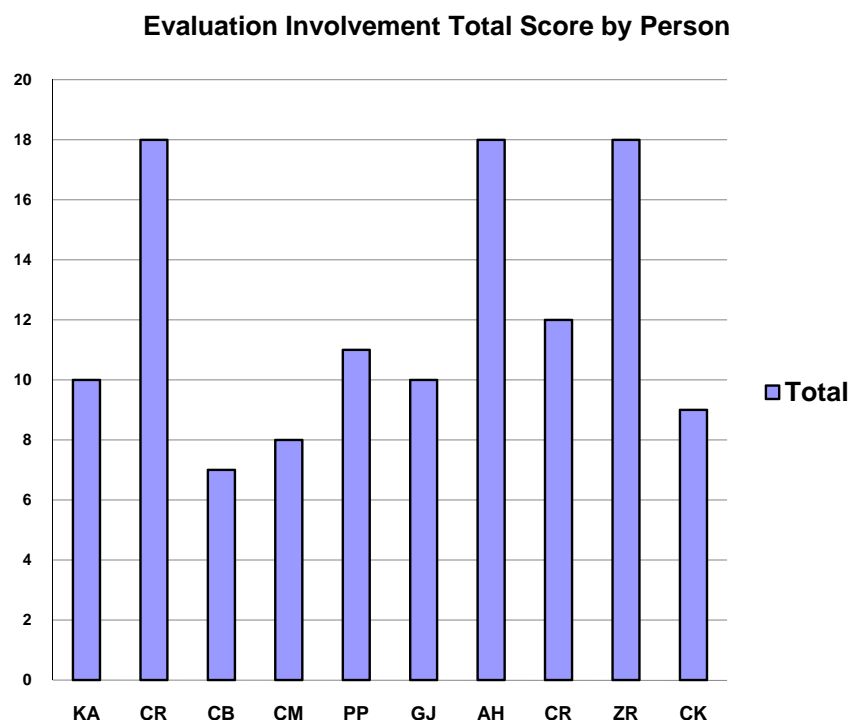
## **EVALUATOR TRAINING, EXPERIENCE AND COMMITMENT TO THE FIELD**

In the second section of the questionnaire, items 3, 4, and 5 are open response items designed to gather information regarding the background the evaluators of the evaluations included in this study. Item 3a focuses on the education and training they have undertaken that has influenced their thinking regarding evaluation and research.

This item is specifically concerned with any workshops, courses or units with a research or evaluation component undertaken by evaluators, the percentage of the learning focused on research and evaluation and their perception of how the undertaking has influenced their thinking regarding research and evaluation. Item 3b is concerned with the conferences or seminars they have attended in the past five years. The year of attendance, the association or society responsible for hosting the conference, whether or not they presented any papers and if the respondent played a role in running or organising the conference. The intent of Item 3 is to gain insight into not only the specifics of the evaluator's background, but also gain some insight into their level of commitment regarding the research and evaluation field. It is expected evaluators with a high degree of education and training in the area of research and evaluation, and a high level of personal advocacy of the field (i.e. organising or running relevant conferences) will perceive the evaluator's role in the evaluation process different to an evaluator with less research and evaluation relevant education and less advocacy or commitment to the field. Furthermore, the influence of these evaluator characteristics on the use of the evaluation information is considered later in this study. The logic underpinning the careful consideration of evaluator background concurs with that of Lake (2005). In his consideration of the characteristics of a 'black belt' evaluator, Lake inspires the notion of evaluators with varying degrees of mastery. Further details regarding the logic underpinning this item have been included in the previous chapter of this dissertation.

Item 4 is concerned with gaining insight into the practice of the evaluator by asking them to recall one research study they have been involved with in the past five years. The six sections of the item are concerned with: the purpose of the study; how the information to be collected was prioritised and what the study was to achieve; the methods used to gather the study information; their personal involvement in terms of the role they played in the research process; the length of the study; and the length of time they were personally involved in the study. Item 5, focusing on one evaluation study they have been involved with in the past five years, follows a similar line of questioning with the addition of two sections. The two additional sections focus on the program evaluated, and extent to which the evaluation included any consideration of the theory or logic underpinning the program: a particular area of interest for this study. Items focusing separately on evaluators' social research background and their evaluation background have been included to address a concern that focusing solely on the gathering of information regarding an evaluator's evaluation background may not have made evident the evaluator's related social research skills and background.

**Table 5.15: Evaluation Involvement Total Score by Person**



Item 6 focuses on gaining insight into the evaluators' perceptions and involvement in the evaluation. The open response line of questioning is identical to that of Item 5 summarised above. As with the other open response items of this interview schedule identified above, the information gathered in response to this item is considered later in this study.

Two instruments were developed for this interview schedule to gain insight into evaluators' perceptions regarding preferred evaluation practice. The first instrument focuses on an evaluator's perceptions of their preferred evaluation practice by asking them to rank five roles an evaluator might take on when undertaking an evaluation study. The second instrument has been developed to gain some insight into the thinking of evaluator's regarding the same five roles, by asking them to rate two facets of each role: evaluator control of the evaluation process and ongoing communication with stakeholders. The two instruments are elaborated on the following sections.

#### Evaluator Perception of Evaluator Practice: Instrument 1

For the purpose of this study, this instrument was developed to gather information regarding the relative importance the evaluators of the evaluation studies place on the various roles an evaluator plays in the course of undertaking an evaluation. The instrument has been included as Item 7 of this interview schedule. Respondents were asked to rate five roles (A Judge of Quality, A Researcher, An Educator, A Facilitator, A Learning Partner) an evaluator might adopt, each varying in terms of evaluator control of the evaluation process and stakeholder involvement in the evaluation process. The intent of this item is to gain insight the evaluators' perception of evaluator practice. It is expected the perception of the evaluators responsible for undertaking the evaluation studies included in this investigation regarding generic evaluator practice, will influence the evaluation process and, in turn, the use of the evaluation information.

The choice of categories for evaluators to rank has been developed based on a review of relevant literature. Many of those involved in laying the foundations

for program evaluation as a discipline, leaned heavily upon traditional secessionist quantitative scientific research methods in an endeavour to establish program evaluation as a new science for assessing program worthiness. In fact, the role of the evaluator advocated by many early theorists is generally a passive one in which they present the information gathered in an evaluation to those responsible for making the decisions (Campbell, 1971; Chen, 1990; Pawson and Tilley, 1994). In fact, in Scriven's (1971) view, a final responsibility of an evaluator is to summarise the mass of data collected in evaluation into a final evaluative judgment. However, other early theorists, such as Joseph Wholey and Robert Stake (Shadish et. al., 1991) felt the evaluator should adopt a more active role with a view to encouraging use of the evaluation information.

As the role of evaluation evolved, so too did the role of the evaluator. In the 1980s, in particular, the evaluator's role was reconsidered and reconceptualised by many to include educator qualities (Cronbach et al., 1980; Mathison, 1994). Further changes to program evaluation were brought about in the late 1980s and 1990s by interest in equity, social justice and community. Guba and Lincoln (1981) advocate the evaluator and program stakeholders as participants in the undertaking of a program's evaluation. Jennifer Greene (1988) perceives the evaluator as being responsible for not only taking into account stakeholder views, but also for creating an environment within which they work collaboratively with stakeholders to define the evaluation's focus, process and outcomes. Similarly, the aim of Fetterman's (1994) empowerment evaluation is to foster self-determination through empowerment and "meaningful" involvement of stakeholders.

Evaluators are asked to weight the five evaluator roles for this item along a continuum from most passive (Judge of Quality and Researcher) to most active (Facilitator and Learning Partner) in terms of involvement with stakeholders. The more passive evaluator roles are least likely to advocate stakeholder involvement in the evaluation process, while the most active are more likely to advocate stakeholder involvement in the evaluation process. Alternatively, the passive Researcher and Judge of Quality roles have greater

control of the evaluation process unlike the roles incorporating more stakeholder involvement. To ask an evaluator to choose only one evaluator role would be too one dimensional, as most evaluators are called upon to play a variety of roles in the conduct of evaluations, as the situation calls for. The logic underpinning the request for evaluators to rank the roles is that their weighting will give insight into their preferences. It is expected these preferences guide their practice.

#### EVALUATOR PERCEPTION OF EVALUATOR PRACTICE: INSTRUMENT 2

A second instrument was developed for this study to gain insight into the evaluators' perception of the evaluator practice. It has been included as Item 8 of the Evaluator Characteristics Interview Schedule I. The instrument asks the respondents to rate on a five-point scale the evaluation practices of *Ongoing Communication with Stakeholders* and *Evaluator Control of the Evaluation Process* with regards to each of the five evaluator roles identified in Item 7. The intent of this instrument is to gain further insight into the evaluators' perception of the roles and their perception of evaluator control inherent in each.

The previous instrument requested evaluators to rank their scores with regards to the five roles. The present instrument has been designed to gain insight into evaluator's perceptions regarding the five roles in terms of communication with stakeholders and evaluator control of the evaluation process. Interpreting the evaluator's ranking of the roles is difficult unless there is some insight into how they perceive the roles. Although there are many facets of the roles that could have been included in the development of this instrument, stakeholder communication and evaluator control have been prioritised as they represent two primary forces in play in the evaluation process.



## EVALUATOR PERCEPTION OF EVALUATOR PRACTICE SUB-SCALES

Thirteen items have been included in this interview schedule to gain insight into: the extent to which evaluators feel the program stakeholders' ideas and opinions should be taken into account by the program evaluator in the evaluation process (*Evaluator Perception of Evaluator Practice Sub-scale 1*, items 9, 10, 11, and 21); the extent to which evaluators feel the context of the program should be taken into account by the program evaluator in the undertaking of the evaluation (*Evaluator Perception of Evaluator Practice Sub-scale 2*, items 13, 14, 15, and 16); and, the extent to which evaluators are of the opinion the program theory, logic or rationale should be taken into account by the program evaluator in the undertaking of the evaluation (*Evaluator Perception of Evaluator Practice Sub-scale 3*, items 17, 18, 19 and 20). These items were intended to form the *Evaluator Perception of Evaluator Practice Sub-scales*. The reliability analyses of the sub-scales are summarised below.

Items 9 to 21 together form the *Evaluator Perception of Evaluators' Practice Scale* of the Evaluator Characteristics Interview Schedule I. Items of this scale focus on: the extent to which evaluators feel the program stakeholders' ideas and opinions should be taken into account by the program evaluator in the evaluation process (items 9, 10, 11, and 21); the extent to which evaluators feel the context of the program should be taken into account by the program evaluator in the undertaking of the evaluation (items 13, 14, 15, and 16); and the extent to which evaluators consider the program theory, logic or rationale should be taken into account by the program evaluator in the undertaking of the evaluation (items 17, 18, 19 and 20). As with the previous two instruments developed for this interview schedule, it is expected this insight into the preferred practices of the evaluator will give some insight into their own practice, in turn influencing the evaluation process.

As mentioned previously in this paper, numerous authors advocate access to and use of the opinions of program stakeholders in the undertaking of the evaluation (e.g. Cummings, 1997; Greene, 1988; Pawson and Tilley, 1997;

Preskill and Torres, 1999a). Similarly, a number of authors have considered the merits of considering the program context in the undertaking of an evaluation (e.g. Hudson-Mabbs, 1993; Braskamp et al., 1982; Shapiro and Blackwell, 1987). Generally, the consensus is consideration of the contextual influences surrounding an evaluand, or the program evaluated, in the evaluation process is likely to yield evaluation information of increased usefulness to the program stakeholders. Finally, the consideration of program theory or logic in the evaluation is an area of particular interest to this study. A number of authors advocate the consideration of program theory or logic in the undertaking of an evaluation (e.g. Chen, 1990; Funnell, 1997; Lipsey and Pollard, 1989; Rogers et al., 2000; Weiss, 1997a; Weiss, 1997b).

Adopted from the post-study interview schedule of Cummings (1997) focusing on the use of evaluation information as perceived by program stakeholders, the items of this section (Item 22 a-g) contribute to a scale included to gain insight into evaluators' perceptions of stakeholder involvement in a range of activities linked to the evaluation process. It is expected that evaluators with a predilection towards stakeholder involvement in the evaluation process will be more likely to conduct evaluations in accordance with their thinking. Cummings found stakeholder involvement to be directly influential on perceived actual use of evaluation information. The items focus on stakeholder involvement in the evaluation conceptualisation in determining the evaluation focus (Item 22a), program theory or logic elaboration (Item 22b), evaluation objectives (Item 22c), target information (Item 22d), information sources (Item 22f), manner of data collection (Item 22e), and reporting of the evaluation information (Item 22g). Item 22 a-g form the *Evaluator Perception of Stakeholder Involvement in the Evaluation Process Scale*.

#### EVALUATOR PERCEPTION OF EVALUATORS' PRACTICE SUB-SCALE 1: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the five items listed in Table 5.16 below, to determine the perception of those responsible for undertaking the evaluations included in this study, regarding the extent to which evaluators should take account of program stakeholders' ideas and opinions in undertaking the evaluation generic evaluator practice. The Cronbach's Alpha value of all five items is  $\alpha = 0.61$ . In the reliability analysis two items were deleted from the scale. Items 10 and 21 were deleted due to low inter-item correlation values and low item total values ( $r_{it} \leq 0.3$ ). A review of inter-item and item-total correlations of the remaining three items indicated all are sound inclusions in the scale.

The alpha value of the final *Evaluator Perception of Evaluator Practice Scale* is  $\alpha = 0.71$ . Inter-item correlations ( $r_{ii'}$ ) for the final scale items ranged from 0.40 to 0.54. All items correlate ( $r_{ii'} \geq 0.3$ ) with two items. The final scale of three items has a mean of 13.44 and a standard deviation of s.d.= 1.51. A summary of the reliability analysis is presented in Table 5.16.

#### EVALUATOR PERCEPTION OF EVALUATORS' PRACTICE SUB-SCALE 2: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the four items listed in Table 5.16 below to determine the perception of those responsible for undertaking the evaluations included in this study, regarding the extent to which evaluators feel the context of the program should be taken into account by the program evaluator in the undertaking of the evaluation. The initial Cronbach's Alpha calculated was  $\alpha = 0.67$ . Item 13 was deleted due to low inter-item correlation values and a low item-total value ( $r_{it} \leq 0.3$ ). A review of inter-item and item-total correlations of the remaining three items indicated all are sound inclusions in the scale.

The alpha value of the final *Evaluator Perception of Evaluator Practice Sub-scale 2* is  $\alpha = 0.82$ . Inter-item correlations ( $r_{ii'}$ ) for the final scale items ranged from 0.40 to 0.91. All items correlate ( $r_{ii'} \geq 0.3$ ) with three items. All three items of the final scale have good item-total values ( $r_{it} \geq 0.3$ ). The final scale of three items has a mean of 12.78 and a standard deviation of s.d.= 1.7. A summary of the reliability analysis is presented in Table 5.16.

#### EVALUATOR PERCEPTION OF EVALUATORS' PRACTICE SUB-SCALE 3: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the four items listed in Table 5.16 below to determine the perception of those responsible for undertaking the evaluations included in this study, regarding the extent to which they are of the opinion that the program theory, logic or rationale should be taken into account by the program evaluator in the undertaking of the evaluation. The Cronbach's Alpha value of all four items is  $\alpha = 0.71$ . No items were deleted from the scale. A review of inter-item and item-total correlations of the remaining three items indicated all are sound inclusions in the scale.

The alpha value of the final *Evaluator Perception of Evaluator Practice Sub-scale 2* is  $\alpha = 0.71$ . Inter-item correlations ( $r_{ii'}$ ) for the final scale items ranged from 0.01 to 0.81. All items correlate ( $r_{ii'} \geq 0.3$ ) with three items. The final scale of four items has a mean of 16.90 and a standard deviation of s.d.= 2.5. A summary of the reliability analysis is presented in Table 5.16.

#### EVALUATOR PERCEPTION OF STAKEHOLDER INVOLVEMENT SCALE

Seven items have been included in the interview schedule to gain insight into evaluators' perceptions of stakeholder involvement in the evaluation process. These seven items were intended to form the *Evaluator Perception of Stakeholder Involvement in the Evaluation Process Scale*. The reliability analysis of the scale is summarised below.

**Table 5.16: Evaluator Perception of Evaluator Practice Scale. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Evaluator Perception of Evaluator Practice Sub-scale 1-</b> Take account of the program stakeholders' ideas and opinions in undertaking the evaluation.			
ITEM 9	Take account of stakeholders' views of the program (e.g. able to put themselves in the stakeholder's shoes)?	2	.54
ITEM 10	Work to establish communication pathways with stakeholders? <b>Item (2<sup>nd</sup>) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.17</math>). When item deleted <math>\alpha = 0.71</math>.</b>	2 <sup>3</sup>	
ITEM 11	Be open to negative feedback about the evaluation process?	2	.61
ITEM 12	Be flexible in undertaking an evaluation?	2	.52
ITEM 21	When planning the evaluation an evaluator should integrate into the evaluation the ideas stakeholders have regarding the evaluation? <b>Item (1<sup>st</sup>) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.15</math>). When item deleted <math>\alpha = 0.65</math>.</b>	1 <sup>3</sup>	
<b>Evaluator Perception of Evaluator Practice Sub-scale 2-</b> Account for Program Context in undertaking the evaluation.			
ITEM 13	Be aware of the history influencing the program being evaluated? <b>Item deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = -0.15</math>). When item deleted <math>\alpha = 0.82</math>.</b>	0 <sup>3</sup>	
ITEM 14	Be aware of the influence of the environment within which the program is set (e.g. relevant policy, availability of resources, level of stakeholder authority...)?	2	.45
ITEM 15	Be aware of the influence of program providers/ practitioners on the program being evaluated?	2	.81
ITEM 16	Be aware of the influence of program clients/ recipients on the program being evaluated?	2	.85
<b>Evaluator Perception of Evaluator Practice Sub-scale 3-</b> Take into account the program rationale, logic or theory in undertaking the evaluation.			
ITEM 17	Understand the rationale, logic or theory of the program?	2	.58
ITEM 18	Take into account the rationale, logic or theory of the program when undertaking the evaluation?	1	.52
ITEM 19	Encourage stakeholders to explore the rationale, logic or theory guiding the program?	2	.57
ITEM 20	Take into account the stakeholders' thinking regarding the rationale, logic, or theory guiding the program?	1	.46
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ) in first run of scale with all items included.			

# EVALUATOR PERCEPTION OF STAKEHOLDER INVOLVEMENT IN THE EVALUATION PROCESS: SCALE ITEM ANALYSIS

Data analysis began with the inclusion of the seven items listed in Table 5.17 below to determine the perception of those responsible for undertaking the evaluations included in this study, regarding the involvement of the program stakeholders in the evaluation process. The alpha value of the final *Evaluator Perception of Stakeholder Involvement Scale* is  $\alpha = 0.81$ . A review of inter-item and item-total correlations of all the items indicated all seven were sound inclusions in the scale. Inter-item correlations ( $r_{ii'}$ ) for items ranged from  $-0.13$  to  $0.71$ . Only items 22c and 22e have an  $r_{ii'} \geq 0.3$  with three items. All other items have an  $r_{ii'} \geq 0.3$  with five items or more. The final scale of seven items has a mean of 25.58 and a standard deviation of s.d.= 4.26. A summary of the reliability analysis is presented in Table 5.17.

**Table 5.17: Perceived Characteristics of Evaluator Regarding Stakeholder Involvement in the Generic Evaluation Process. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>To what extent, in your opinion, should stakeholders be involved in each of the following areas:</b>			
ITEM 22a	Defining what the study focuses on?	5	.63
ITEM 22b	Exploring the logic/theory or rationale underpinning the program?	6	.75
ITEM 22c	Developing the objectives of the study?	3	.46
ITEM 22d	Identifying what information is to be collected?	5	.66
ITEM 22e	Deciding how the information is to be collected?	3	.36
ITEM 22f	Identifying whom the information will be collected from?	5	.59
ITEM 22g	Determining how the information will be reported?	5	.57
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.			

## SUMMARY

This section of the chapter has detailed the content, focus and logic underpinning the five sections of the Evaluator Interview Schedule I. The first section concentrates on characteristics of the involvement of the evaluator in the evaluation, while the second section focuses on the evaluators' evaluation/research background in terms of training, education, experience, professional conference/seminar involvement or contribution to the field, details of past research and evaluations study involvement, and insight into the current evaluation process. The third section details two instruments developed for this interview schedule to gather information from evaluators regarding their perception of preferred evaluation practice. Finally, the fifth section focuses on the items included in the two scales included in the interview schedule. The first scale is concerned with gaining insight into evaluator's thinking regarding evaluator practice with regards to a variety of evaluation activities. The second has been included to gain insight into evaluator's thinking regarding stakeholder involvement in a variety of evaluation processes.

Data gathered in the administration of the Evaluator Characteristics Interview Schedule was used in the analysis of the scale and sub-scale items. A number of the items are open response format; therefore the results of these items are reported later in this study. Similarly, the results of the four scales developed for this study concerned with evaluator practice are discussed later. The intent of this section was to verify the four scales built for inclusion in this interview schedule, the *Evaluator Perception of Evaluator Practice Sub-scales* and the *Evaluator Perception of Stakeholder Involvement in the Evaluation Process Scale*. Interview data was used for item analysis. Inter-item correlations were calculated to determine the fit of the items of each scale and sub-scale. Two items were dropped from the *Evaluator Perception of Evaluator Practice Sub-scale 1* due to low inter-item correlations ( $r_{ii} \leq 0.3$ ) and/or low inter-total correlations ( $r_{it} \leq 0.3$ ). One item was dropped from the *Evaluator Perception of Evaluator Practice Sub-scale 2* due to low inter-item correlations ( $r_{ii} \leq 0.3$ )

and/or low inter-total correlations ( $r_{it} \leq 0.3$ ). All items of the *Evaluator Perception of Evaluator Practice Sub-scale 3* and the *Evaluator Perception of Stakeholder Involvement in the Evaluation Process Scale* were retained. The Cronbach's Alphas for the four scales and sub-scales were graded ( $\alpha \leq 0.7$ ).

#### PROCESS USE INTERVIEW SCHEDULE

Michael Patton in the preparation of his third edition of *Utilisation-Focused Evaluation* (1998) became aware of “myopia” (p. 288; Patton 2004) in his narrow conceptualisations of evaluation use in previous work. He felt this narrow conceptualisation of use resulted in the disregard of the influence of the evaluation process on the people and organisations involved. In response, he defined ‘process use’ (Patton 1997, 1998) as changes in the thinking and/or behaviour of those involved in the evaluation, resulting from the learning that takes place in the evaluation process, and changes to program or organisational culture and processes.

The Process Use Interview Schedule developed for this research is concerned with gathering information from stakeholders of an evaluation study, regarding their level of involvement in the evaluation study. The survey also focuses on the extent to which their involvement in the evaluation process has influenced: their thinking regarding the program or issues beyond the program, any decisions they were involved in making, and their practice. The reliability analysis of the *Perceived Process Use Scale* (Table 5.18), the only scale included in this interview schedule, is detailed below.

The first portion of the Process Use Interview Schedule developed for this research is concerned with gathering information from stakeholders of each evaluation study regarding their level of involvement in the evaluation study. In considering the amount of involvement with the evaluation study stakeholders had, it must first be determined if they have been involved in the evaluation and in what ways. Five open response items have been included



focused on: the contact they have had with the evaluators (Item 1); any discussions they have had with evaluators in which they have provided program information (Item 2); any discussions they have had with evaluators in which they have provided input into the planning or management of the evaluation (Item 3); any involvement they have had with reference groups or steering committees concerned with the evaluation (Item 4); and any other evaluation activities they have been involved with (Item 5).

The remaining items of the interview schedule are concerned with stakeholders' perception of the influence of any involvement in the evaluation on their; thinking (items 6 and 7), thinking about issues other than the program (Items 8 and 9), any decision they were involved in making (Items 10 and 11), and their practice (items 12 and 13). These types of influence have been targeted as they have been identified by numerous authors in typologies of evaluation influence (e.g. Alkin et al., 1985; Cousins and Leithwood, 1986, 1993; Leviton and Hughes, 1981; Shulha and Cousins, 1997).

The first item of each pair is an open response item to enable them to contemplate each type of influence and to respond fully. The second item of the pair then asks them to rate the extent of each type of influence on the one five point scale. The four scaled response items 7, 9, 11, and 13 together form the *Process Use Scale* developed for the study. The final item of the survey (Item 14) seeks further comments regarding the influence of stakeholder involvement.

## **PROCESS USE SCALE**

Four quantitative items have been included in the Perceived Process Use interview schedule to gain insight into stakeholder thinking, regarding the extent to which their involvement in the evaluation process has influenced their thinking, practice and decision-making. These four items together were intended to form the *Process Use Scale* of this study. The reliability analysis of the scale is summarised below.

## PERCEIVED PROCESS USE: SCALE ITEM ANALYSIS

The item analysis began with the inclusion of the four items (Item 7, Item 9, Item 11 and Item 12) listed in the table below, to determine the extent of influence on stakeholder thinking, practice and decision-making caused by stakeholder involvement in the evaluation (Patton, 1998). Cronbach's Alpha of all four items was  $\alpha = 0.83$ . In the reliability analysis, all four items were found to be good inclusions to the scale. Inter-item correlations ( $r_{ii'}$ ) for items range from 0.45 to 0.71. All items have a  $r_{ii'} \geq 0.3$  with three other scale items. The item-total values for all scale items are good ( $r_{it} \geq 0.6$ ). The final scale of four items has a mean of 6.38 and a standard deviation of s.d.= 3.84. A summary of the reliability analysis is shown in Table 5.18.

## SUMMARY

The preceding section detailed the items of the Process Use Interview Schedule. The interview schedule contains a mix of open response and scaled response items, focusing first on stakeholders' perception of their involvement in the evaluation, and then on the extent to which this involvement has influenced their thinking regarding the program and issues beyond the program, their practice and any decisions they were involved in making. An analysis of The *Process Use Scale* comprised of items 7, 9, 11, and 13 is undertaken. The *Perceived Process Use Scale* detailed above represents the characteristics representing process use included in this study. Interview data was used for analysis. With regards to the scale, no items have been deleted due to a weak inter-item correlation ( $r_{ii'} \leq 0.03$ ). The scale has been verified (Cronbach's Alpha  $\geq 0.7$ ). The analysis of the scale helped confirm what was anticipated in terms of scale and sub-scale content and focus.

**Table 5.18: Perceived Process Use Scale. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
ITEM 7	To what extent do you think this involvement affected your thinking about the program?	3	0.70
ITEM 9	To what extent do you think your involvement affected your issues other than the program?	3	0.76
ITEM 11	To what extent do you think your involvement influenced any decisions you were involved in making?	3	0.62
ITEM 13	To what extent do you think your involvement influenced your practice?	3	0.66
Note:	<sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.		

## STAKEHOLDER INTERVIEW SCHEDULE II

The Stakeholder Interview Schedule II was developed to gather information from the stakeholders of the evaluations that have been included in this longitudinal study, focusing on the characteristics of the evaluation studies and their influence on the use of the evaluation information. This interview schedule was administered post-evaluation study following the release of the final reports to the stakeholders. The data collected in the administration of the Stakeholder Interview Schedule II has been used below to analyse the scale items.

The items of the interview schedule focus on stakeholders' perspectives of: the evaluation study characteristics; the influence of the evaluation information; the characteristics of the evaluators responsible for undertaking the evaluation studies; the characteristics of the environment influencing the evaluation process; their personal involvement in the evaluation studies; the use of the program's theory, logic or rationale in the evaluation; the influence of their involvement in any work to understand or elaborate the program theory, logic

or rationale; and their use of the final evaluation reports and the evaluation information.

The sections of the interview schedule, the items and the scales developed to analyse the information collected, are detailed more fully below. All scale items, with the exception of the items in the final section of the interview schedule focusing on use of the evaluation information and Item 12, require responses on a five-point scale. In addition, a number of open response items have been included with a view to gaining further insight into the areas of focus. This information provides further insight into the quantitative findings of the scale items.

The quantitative items of the interview schedule focus on stakeholders' perspectives of: the evaluation study characteristics; the influence of the evaluation information; the characteristics of the evaluators responsible for undertaking the evaluation studies; the characteristics of the environment influencing the evaluation process; their personal involvement in the evaluation studies; the use of the program's theory, logic or rationale in the evaluation; the influence of their involvement in any work to understand or elaborate the program theory, logic or rationale; and their use of the final evaluation reports and the evaluation information. The logic underpinning the inclusion of each scale and the focus of the items is detailed in Chapter V of this dissertation. The reliability analysis of each of the eight scales of this interview schedule is detailed in the following sections.

The initial section of the Stakeholder Interview Schedule II begins with an open response item focused on the stakeholders' perceptions of the main purpose(s) of the evaluation study. This item corresponds with Item 16 of the Stakeholder Interview Schedule I and is intended to highlight differences in stakeholder perceptions before and after the evaluation study process.

## STAKEHOLDER PERSPECTIVE OF EVALUATION STUDY CHARACTERISTICS SCALE

Nineteen items of the interview schedule were developed to gain insight into stakeholders' perspectives regarding the characteristics of the evaluation study. These items were intended to form the *Stakeholder Perspective of Evaluation Study Characteristics Scale*. The reliability analysis of the scale, including two sub-scales, is considered in the following section.

The first scale focuses on stakeholder perceptions regarding characteristics of the evaluation study. The second is intended to gather information from stakeholders regarding their perceptions of the influence of the evaluation study. The items of each sub-section are detailed below. Items of this section contribute to the *Stakeholder Perspective of Evaluation Study Characteristics Scale* composed of two sub-scales: *The Stakeholder Perceived Characteristics of Commitment to Evaluation Study- Post Sub-scale* and the *Stakeholder Perceived Characteristics of the Evaluation Study Sub-scale*.

Item 2 (evaluation need), Item 3 (usefulness of the evaluation information), Item 5 (usefulness of the evaluation information to the organisation responsible for program delivery), Item 6 (appropriate for the stage of the program's life), and Item 12 (the extent to which the stakeholder supports the evaluation study) have been included in the Stakeholder I Interview Schedule and have been included in the *Stakeholder Perceived Characteristics of Commitment to Evaluation Study- Post Sub-scale* of this scale. The inclusion of these items both pre- and post- study, with minor changes to item wording in terms of tense, allows for comparison and contrast of stakeholder expectations and perspectives prior to the conduct of the evaluation study with their perspectives of the study following dissemination of the final report.

Item 12 focusing on stakeholder support for the program requires respondents to choose between two options regarding their support or opposition to the program. Item 13 then asks them to rate on the five-point scale, the strength of that choice which has been coded as preponderance.

A second sub-scale included in the section, *Stakeholder Perceived Characteristics of the Evaluation Study Sub-scale* focuses on stakeholder perceptions of characteristics of the evaluation study including: the quality of the data collected in the study (Item 4); the study timeliness for any pending decisions (Item 7); the extent to which the study findings are politically acceptable (Item 8); the agreement of the study findings with other sources of information (Item 9); the influence of their role in their use of the study information (Item 10); the appropriateness of the evaluation study methods (Item 11); the extent to which the study findings are supportive of the program (Item 14); the extent to which the study findings are critical of the program (Item 16); the extent to which they agree with the study findings (Item 18); and the extent to which they believe the study findings are relevant (Item 29).

Four open response items have been included to gain further insight regarding the ways in which stakeholders feel the study findings are supportive (Item 15) and critical (Item 17) of the program, how the findings agree with their own personal assessment of the program (Item 19), and how they feel the evaluation focused on relevant issues (Item 30).

#### STAKEHOLDER PERSPECTIVE OF EVALUATION STUDY CHARACTERISTICS: SCALE ITEM ANALYSIS

*Sub-scale 1: Stakeholder Perceived Characteristics of Commitment to Evaluation Study- Post.* The five items of this sub-scale have been included in the Stakeholder Interview Schedule I with a view to comparing and contrasting stakeholders pre-evaluation study expectations and perspectives with their post-evaluation study perspectives. An analysis of this sub-scale has not been undertaken here, as the scale has been previously analysed (see Table 5.7) as the *Perceived Characteristics of Stakeholder Commitment to the Evaluation Study Scale- Pre* of the Stakeholder Interview Schedule I.

*Sub-scale 2: Stakeholder Perceived Characteristics of the Evaluation Study.* The initial Cronbach's Alpha calculated was  $\alpha = 0.79$ . Items 8 and 16 were

deleted from the sub-scale due to low inter-item correlation values ( $r_{ii'} \geq 0.3$  with only one other item) and a low item-total correlation value ( $r_{it} = 0.21$  and  $r_{it} = 0.20$ , respectively). The deletion of these items increased Cronbach's Alpha of the scale to  $\alpha = 0.83$ . Item 29 of the sub-scale correlates with three other items ( $r_{ii'} \geq 0.3$ ), while all other sub-scale items correlate ( $r_{ii'} \geq 0.3$ ) with four or more sub-scale items. The inter-item correlation ( $r_{ii'}$ ) values for the final eight sub-scale items range from 0.08 to 0.67. The sub-scale of eight items has a mean of 27.34 and a standard deviation of s.d.= 5.78. The reliability analysis is summarised in Table 5.19.

#### **STAKEHOLDERS' PERSPECTIVE OF THE INFLUENCE OF THE EVALUATION STUDY FINDINGS SCALE**

The interview schedule includes five items to gain insight into the stakeholders' perspectives regarding the influence the evaluation study findings have had on them. The 5 scaled response items focus on the influence of the findings on; their thinking regarding the program (Item 20), their thinking regarding issues beyond the program (Item 21), any program decisions (Item 23), decisions beyond the immediate program (Item 25), and support for views or positions about the program which people already held (Item 27). These five items have been included to form the *Stakeholder's Perspective of the Influence of the Evaluation Study Findings Scale*. Each scaled response item is followed by an open response item (items 22, 24, 26, 28) developed to gain further insight regarding each area of influence from the perspective of the stakeholder.

#### **STAKEHOLDERS' PERSPECTIVE OF THE INFLUENCE OF THE EVALUATION STUDY FINDINGS SCALE: SCALE ITEM ANALYSIS.**

Data analysis of the *Stakeholders' Perspective of the Influence of the Evaluation Study Findings Scale* began with the inclusion of the 5 items listed in Table 5.20. Cronbach's Alpha of all five items was  $\alpha = 0.70$ , but with the deletion of Item 27 in the reliability analysis, the scale Alpha was increased to  $\alpha = 0.76$ . Item 27 was deleted from the scale, as it did not correlate with any other scale items ( $r_{ii'} \geq 0.3$ ) and had a very low item-total correlation value ( $r_{it} =$

0.21). Inter-item correlations ( $r_{ii'}$ ) for items range from 0.29 to 0.63. All final scale items have an  $r_{ii'} \geq 0.3$  with the two other scale items or more. The final scale of four items has a mean of 9.71 and a standard deviation of s.d.= 3.43.

**Table 5.19: Stakeholder Perspective of the Evaluation Study Characteristics Scale. Item Analysis Results.**

Scale Items Grouped by Area of Focus		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Sub-scale 1: Stakeholder Perceived Characteristics of Commitment to Evaluation Study- Post<sup>3</sup></b>			
ITEM 2	Do you feel this evaluation study was needed?		
ITEM 3	Do you think this study provided information useful to the program?		
ITEM 5	Do you think this study provided information useful to _____(the organisation responsible for the program)?		
ITEM 6	Was the evaluation appropriate for the stage of the program's life?		
ITEM 12/13	How strongly do you support this study? <sup>4</sup>		
<b>Sub-scale 2: Stakeholder Perceived Characteristics of the Evaluation Study</b>			
ITEM 4	Do you feel the quality of the data that has been collected in the study is sound?	7	0.73
ITEM 7	Were the evaluation findings timely for any pending decisions?	5	0.53
ITEM 8	Were the implications of the study politically acceptable? <b>Item (1<sup>st</sup>) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.21</math>). When item deleted <math>\alpha = 0.81</math>.</b>	1 <sup>5</sup>	
ITEM 9	Were other sources of information in agreement with the study findings?	4	0.51
ITEM 10	Did your organisational role influence the way you used the evaluation information?	4	0.42
ITEM 11	Do you feel the evaluation study method was appropriate?	6	0.68
ITEM 14	To what extent do you believe the findings of the study are supportive of the program?	4	0.45
ITEM 16	To what extent do you believe the findings of the study are critical of the program? <b>Item (2<sup>nd</sup>) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.20</math>). When item deleted <math>\alpha = 0.83</math>.</b>	1 <sup>4</sup>	
ITEM 18	Do the study findings agree with your own personal assessment of the program?	4	0.59
ITEM 29	To what extent do you feel the evaluation focused on issues considered relevant by those interested in the program?	4	0.43
<p>Note: <sup>1</sup> Represents the number of other scale items the item correlates with (<math>r_{ii'} \geq 0.3</math>).</p> <p><sup>2</sup> Represents the item-total correlation value (corrected) for the item with the final scale.</p> <p><sup>3</sup> The analysis of this scale has been undertaken with Stakeholder Interview I data, therefore only the text of the sub-scale items is reported here.</p> <p><sup>4</sup> Item represents a composite of Items 12 and 13.</p> <p><sup>5</sup> Represents the number of other scale items the item correlates with (<math>r_{ii'} \geq 0.3</math>) in first run of scale with all items included.</p>			



## STAKEHOLDERS' PERSPECTIVE OF THE EVALUATION TEAM CHARACTERISTICS

The second section of the interview schedule focuses on stakeholder perceptions of the evaluation team and of the evaluation environment. Numerous authors have found characteristics of the evaluator to have influence on the use of the information from an evaluation study (items 34 and 47) (e.g. Greene, 1988; Hudson-Mabbs, 1993; Lake, 2005; Seigel and Tuckel, 1985). In particular, the stakeholders' perception of the evaluators has been found to have influence when there is potential for dissension among the stakeholders of the evaluation regarding the findings an evaluation study and when the findings are to be disseminated outside the program (Hudson-Mabbs, 1993; Leviton and Hughes, 1981).

**Table 5.20: Stakeholders' Perspective of the Influence of the Evaluation Study Findings Scale. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
ITEM 20	To what extent did the evaluation information influence your thinking about the program?	3	0.56
ITEM 21	To what extent did the evaluation information influence your thinking about issues beyond the program?	2	0.54
ITEM 23	To what extent did the evaluation have an impact on any program decisions?	2	0.53
ITEM 27	To what extent do you think the evaluation information was used as support for views or positions about the program which people already held? <b>Item deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) of item low (<math>r_{it} = 0.21</math>). When item deleted <math>\alpha = 0.77</math>.</b>	0 <sup>3</sup>	
Note:		<sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents number of items $r_{ii'} \geq 0.3$ in first run of scale with all items included.	

Seven items of the interview schedule have been included to gain insight into the stakeholders' perspective of the evaluation team, regarding the extent to which they were competent, tolerant, approachable, protective of their ideas

regarding the evaluation study, endeavoured to develop relationships with stakeholders and listened to stakeholders' ideas regarding the evaluation. These seven items were intended to form the *Stakeholders' Perspective of the Evaluation Team Characteristics Scale*.

The first item of the section (Item 31) is a screening item requiring a dichotomous response with regards to whether the stakeholder has had an opportunity to meet with the evaluators. If not, stakeholders were not asked to respond to the remaining items of the section. If they had met with the evaluators, the next two open response items gained further insight regarding the occasions they had met with the evaluators (Item 32) and why they believed the evaluation team was chosen to conduct the evaluation study (Item 33).

Furthermore, numerous authors advocate access to and use of the opinions and thoughts of stakeholders in the undertaking of the evaluation, with a view to increasing the usefulness of the evaluation information (e.g. Cummings, 1997; Greene, 1988; Pawson and Tilley, 1997; Preskill and Torres, 1999b). Items 35, 36, 37, 38 and 41 have been included to gain insight into stakeholders' thinking regarding the evaluators' openness to listening to the ideas of stakeholders, regarding the evaluation and capacity to incorporate them into the evaluation process. Items requiring responses on a five point scale focusing on the characteristics of the evaluation team, include the extent to which they; were approachable (Item 35), were protective of their ideas for the evaluation (Item 36), endeavoured to develop relationships with stakeholders (Item 37), were open to stakeholders ideas regarding the evaluation (Item 38), were tolerant of change (Item 41), were competent to undertake the evaluation study (Item 34), and competently conducted the evaluation study (Item 47). These seven items have been included to form the *Stakeholders' Perspective of the Evaluation Team Characteristics Scale*.

A combination of open and closed response items focus on the extent to which evaluators felt that; they were given sufficient latitude to undertake the evaluation, and circumstances surrounding the evaluation did not impede its

progress in any way. Open response items were added to the schedule to enable stakeholders to elaborate on their thinking regarding the extent to which evaluators were tolerant of change (Item 42) and the extent to which they were competent at conducting the evaluation (Item 48). The reliability analysis of the scale is summarised below.

#### STAKEHOLDERS' PERSPECTIVE OF THE EVALUATION TEAM CHARACTERISTICS: SCALE ITEM ANALYSIS.

A summary of the reliability analysis is shown in Table 5.21. Cronbach's Alpha of all seven items was  $\alpha = 0.82$ , but with the deletion of Item 36 (this item correlated with only two other items  $r_{ij} \geq 0.3$  and had a very low item-total correlation value ( $r_{it} = 0.25$ ) in the reliability analysis, the scale Alpha was increased to  $\alpha = 0.86$ . Although Item 34 only correlates ( $r_{ij} \geq 0.3$ ) with two other scale items, the deletion of the item would have resulted in only a minor increase in the scale value of Cronbach's Alpha ( $\alpha = 0.86$ ). Furthermore, the item-total correlation value for Item 34 is reasonable ( $r_{it} = 0.42$ ). The item has been retained for the range it adds to the scale.

The deletion of Item 41 would have increased the scale value of Cronbach's Alpha to  $\alpha = 0.87$ . However, the item correlates with three other scale items ( $r_{ij} \geq 0.3$ ), and the item-total correlation value for Item 41 is reasonable ( $r_{it} = 0.44$ ). Therefore, the item has been retained for the range it adds to the scale.

Inter-item correlations for the final six scale items range from 0.20 to 0.80. All final scale items have an  $r_{ij} \geq 0.3$  with the two other scale items. The final scale of six items has a mean of 21.08 and a standard deviation of s.d.= 4.60.

#### **Stakeholders' Perspective of the Evaluation Study Environment Scale**

Six items were developed to gather information from stakeholders regarding their perception of the evaluation environment, with three scaled response items followed by an open response item to enable elaboration on the scale item response. The logic underpinning the items is focused on understanding

any influences constraining the evaluation process. The items focus on stakeholders' perceptions regarding the extent to which the evaluation was able to proceed as planned (Item 39 and Item 40), the extent to which

**Table 5.21: Stakeholders' Perspective of the Evaluation Team Characteristics. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Evaluator Competency</b>			
ITEM 34	To what extent do you think the evaluators chosen were sufficiently competent to undertake the evaluation study?	2	0.42
ITEM 47	Given those circumstances, to what extent were the evaluators competent at conducting the evaluation?	5	0.79
<b>Evaluator Practice</b>			
ITEM 35	To what extent was the evaluation team approachable?	4	0.75
ITEM 36	To what extent was the evaluation team protective of their ideas for the evaluation? <b>Item deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) for item low (<math>r_{ii'} = 0.23</math>). When item deleted <math>\alpha = 0.87</math>.</b>	2 <sup>3</sup>	
ITEM 37	To what extent did the evaluation team endeavor to develop relationships with stakeholders?	4	0.72
ITEM 38	To what extent did the evaluation team listen to the ideas stakeholders had regarding the evaluation, even when they challenged their own?	5	0.86
ITEM 41	To what extent were the evaluators tolerant of change?	3	0.44
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents number of items $r_{ii'} \geq 0.3$ in first run of scale with all items included.			

evaluators were given sufficient latitude in undertaking the evaluation (Item 43 and Item 44), and the extent to which evaluators were free to undertake the evaluation (Item 45 and Item 46). Insight into stakeholders' perception of the environment of the evaluation is expected to illuminate any barriers or obstacles influencing the evaluation process which the evaluators had to overcome in undertaking the evaluation. The three scaled response items (items 39, 43 and 45) have been included to form the *Stakeholder Perspective*

of the *Evaluation Study Environment Scale*. These three items were intended to form the *Stakeholder Perspective of the Evaluation Study Environment Scale*. The reliability analysis of the scale is summarised in the section below.

#### STAKEHOLDERS' PERSPECTIVE OF THE EVALUATION STUDY ENVIRONMENT: SCALE ITEM ANALYSIS.

Data analysis of the *Stakeholders' Perspective of the Evaluation Study Environment Scale* began with the inclusion of the three items focusing on the perspectives of the stakeholders regarding characteristics of the environment of the evaluation studies included in this study. A summary of the reliability analysis is shown in Table 5.22. Cronbach's Alpha of all three items was  $\alpha = 0.44$ , but with the deletion of Item 39 in the reliability analysis the scale Alpha was increased to  $\alpha = 0.60$  and this item correlated with only one other item  $r_{ii'} \geq 0.3$  and had a very low item-total correlation value ( $r_{it} = 0.11$ ). The final two scale items have an inter-item correlation of 0.60. The final scale of two items has a mean of 7.53 and a standard deviation of s.d. = 2.0.

**Table 5.22: Stakeholders' Perspective of the Evaluation Study Environment Scale. Item Analysis Results.**

Scale Items Grouped by Area of Focus.		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
ITEM 39	To what extent were there times when the evaluation process was unable to proceed as planned? <b>Item deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) for item low (<math>r_{ii'} = 0.11</math>). When item deleted <math>\alpha = 0.87</math>.</b>	1 <sup>3</sup>	
ITEM 43	To what extent do you feel the evaluators were given sufficient latitude to exercise sound professional judgment in undertaking the evaluation?	1	0.56
ITEM 45	Considering circumstances surrounding the evaluation, to what extent do you feel the evaluators were free to undertake the evaluation?	1	0.56
Note:	<sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents number of items $r_{ii'} \geq 0.3$ in first run of scale with all items included.		

## STAKEHOLDERS' PERSPECTIVE OF THE EVALUATION STUDY INVOLVEMENT SCALE

The third section of the interview schedule includes ten items requiring a response on a five-point scale focusing on the type and degree of stakeholder involvement in the evaluation process. The items of this section (items 49, 50, 51, 52, 53, 56, 58, and 70) which contribute to the *Stakeholder Perspective of Evaluation Study Involvement Scale*, have also been included in the *Stakeholder Expectation of Involvement in the Evaluation Study Scale* included in the Stakeholder I Interview Schedule (Table 5.8), with minor changes to item wording regarding tense. As detailed in the section focusing on that scale, numerous authors have identified a strong link between stakeholder involvement in the evaluation study and the use of its information (e.g. Cousins and Earl, 1995; Cummings, 1997; Patton, 1997; Preskill and Torres 1997, 2000a). In particular, “social constructivist learning” (Preskill and Torres, 2000a, p. 31) occurs when individuals consider information together, as is the case when stakeholders are involved in the program theory elaboration process (Item 70).

For the purposes of this study, relevant data has been collected both pre-and post-evaluation study conduct with a view to gaining insight into discrepancies between stakeholder expectations, regarding involvement in the evaluation study and the extent to which they were actually involved. Furthermore, the influence of this discrepancy and the influence of study involvement on the influence of the evaluation’s information, will be considered later in this work. The reliability analysis of the scale is summarised earlier in this dissertation as the *Perceived Stakeholder Expectation of Involvement in the Evaluation Study Scale* (Stakeholder Interview Schedule I, Table 5.8).

## STAKEHOLDERS' PERSPECTIVE OF THE EVALUATION STUDY INVOLVEMENT SCALE: SCALE ITEM ANALYSIS.

A scale item analysis of the *Stakeholders' Perspective of the Evaluation Study Involvement Scale* has been undertaken earlier as the *Perceived Stakeholder*

*Expectation of Involvement in the Evaluation Study Scale* included in the Stakeholder Interview Schedule I (Table 5.8). Therefore, further analysis of the scale is not necessary at this point. The text of the eight items included in the *Stakeholders' Perspective of the Evaluation Study Involvement Scale* have been included below in Table 5.23.

#### **STAKEHOLDERS' PERSPECTIVE OF THE INFLUENCE OF THE USE OF PROGRAM THEORY IN THE FINAL EVALUATION REPORT SCALE**

The items of this section were developed to gain insight into stakeholders' perception of use of program theory in the evaluation report and the influence of the program theory in the evaluation on their own thinking and practice. As Cummings et al. (2001) write in their paper, the final description of the program serves to inform stakeholders about the program evaluated and is critical to both the quality of the evaluation study, as well as the manner and degree of the evaluation study information influence. One of the primary intents of this study is to gain insight into the influence of use of the program's theory in undertaking an evaluation study.

The section begins with a screening item to determine the extent to which the evaluation report contained any information about the program theory or logic (Item 59). If the report did not at all, no further items of the section were asked of the stakeholder. If it did, stakeholders were asked a scaled response item about the extent to which the program theory presented made sense to them (Item 60), followed by an open response item focusing on their thoughts regarding the source of the program theory used by the evaluators (Item 61). Finally, two items focusing on the extent to which the stakeholder agrees with the program theory version used (Item 62) and why (Item 63) concludes the interview schedule portion. These items were included to form the *Stakeholders' Perspective of the Influence of the Use of Program Theory in the Final Evaluation Report Scale* to gain insight into the stakeholders' perception of the program theory or logic elaborated in the final evaluation report. The logic of these items is that an understanding of the stakeholder's

perception of the program theory use in the evaluation, is an important precursor to interpreting the influence of the program theory. A combination of closed and open response items then centre on the stakeholder's perception as to the extent to which the program theory use in the evaluation influenced their thinking regarding the program (items 64 and 65), practice (items 66 and 67) and any program decisions made or which they expect to make (items 68 and 69).

**Table 5.23: Stakeholders' Perspective of the Evaluation Study Involvement Scale Items.**

SCALE ITEMS GROUPED BY AREA OF FOCUS <sup>1</sup>	
<b>Expected Involvement (e.g. by org. responsible for delivery, stakeholder org., etc.)?</b>	
ITEM 50	Was your involvement in planning the evaluation study expected (e.g. by org. responsible for delivery, your org. etc.)?
ITEM 52	Was your involvement in the conduct of the evaluation study as expected (e.g. by org. responsible for delivery, your org., etc.)?
<b>Involvement in the Evaluation Process</b>	
ITEM 49	Were you involved in planning the evaluation study?
ITEM 51	Were you involved in the conduct of the evaluation study?
ITEM 70	To what extent were you involved in any work to understand/explain the theory or logic underpinning the program?
ITEM 53	Were you involved in regular communication with the evaluation team?
ITEM 56	Did you provide feedback to the evaluation team on issues and concerns regarding the evaluation?
ITEM 58	Were you involved in developing the recommendations from the study?
<b>Note:</b>	<sup>1</sup> The analysis of this scale has been undertaken with Stakeholder Interview I data, therefore only the text of the sub-scale items is reported here.



Six scaled items (items 59, 60, 62, 65, 67, 69) were intended to form the Stakeholders' Perspective of the Influence of the Use of Program Theory in the Final Evaluation Report Scale. The reliability analysis of the scale is summarised below.

#### STAKEHOLDERS' PERSPECTIVE OF THE INFLUENCE OF THE USE OF PROGRAM THEORY IN THE FINAL EVALAUTION REPORT SCALE: SCALE ITEM ANALYSIS

Initial attempts to analyse the data of the *Stakeholders' Perspective regarding the Influence of the Use of Program Theory in the Final Evaluation Report Scale* put the following items into one sub-scale: Item 59 (Did the evaluation report contain information about the program theory or logic?), Item 60 (To what extent was the program theory presented in a manner that made sense to them?) and Item 62 (To what extent did they agree with the theory used?). Item 65 (Program theory influence on their thinking regarding program), Item 67 (Program theory influence on their practice) and Item 69 (Program theory influence on any decisions regarding the program) were put into a second sub-scale. The logic delineating these sub-scales was that the first three items focus on gaining insight regarding the stakeholders' perception of the program theory used in the evaluation report, while the last three items focus on the stakeholders' perspective of the influence the program theory has had on their thinking, practice, and decisions regarding the program. Although the items of the second sub-scale worked well together, the items of the first did not.

In an attempt to include all items in the scale, the two sub-scales were combined. However, in the reliability analysis the first three items (items 59, 60 and 62) were deleted from the scale due to low Cronbach's Alpha values, low inter-item correlations ( $r_{ii'} \leq 0.3$ ) and low item-total correlations ( $r_{it} \leq 0.30$ ). Cronbach's Alpha of all six items was  $\alpha = 0.64$ , but with the deletion of items 59, 60 and 62 in the reliability analysis, the scale Alpha was increased to  $\alpha = 0.81$ . Inter-item correlations ( $r_{ii'}$ ) for the remaining three scale items range from 0.39 to 0.75. All final scale items have an  $r_{ii'} \geq 0.3$  with two other scale items. The final scale of three items has a mean of 6.50 and a standard

deviation of s.d.= 2.97. A summary of the reliability analysis is shown in Table 5.24.

**Table 5.24: Stakeholders' Perspective of the Influence of the Use of Program Theory in the Final Evaluation Report Scale: Item Analysis Results**

SCALE ITEMS GROUPED BY AREA OF FOCUS		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
ITEM 59	Did the evaluation report contain information about the program theory or logic? <b>Item (2<sup>nd</sup>) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) for item low (<math>r_{it} = 0.21</math>).<sup>4</sup> When item deleted <math>\alpha = 0.73</math>.</b>	1 <sup>3</sup>	
ITEM 60	Was it presented in a manner that made sense? <b>Item (3<sup>rd</sup>) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) for item low (<math>r_{it} = 0.17</math>).<sup>4</sup> When item deleted <math>\alpha = 0.81</math>.</b>	1 <sup>3</sup>	
ITEM 62	To what extent do you agree with the program theory or logic used in the evaluation? <b>Item (1st) deleted due to low <math>r_{ii'}</math> values. Item-total value (corrected) for item low (<math>r_{it} = -0.02</math>).<sup>4</sup> When item deleted <math>\alpha = 0.70</math>.</b>	0 <sup>3</sup>	
ITEM 65	To what extent do you believe the use of the program's theory in the evaluation influenced your thinking about the program?	2	0.50
ITEM 67	To what extent do you believe the use of the program's theory in the evaluation influenced your practice?	2	0.81
ITEM 69	To what extent do you believe the use of the program's theory in the evaluation influenced any decisions you've made or will make about the program?	2	0.66
Note: <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Represents number of items $r_{ii'} \geq 0.3$ in first run of scale with all items included. <sup>4</sup> Represents item-total (corrected) prior to deletion from scale.			

#### STAKEHOLDERS' PERSPECTIVE OF THE INFLUENCE OF THEIR INVOLVEMENT IN THE PROGRAM THEORY ELABORATION PROCESS SCALE

The items of this section were developed to gain insight into stakeholders' perspective of the influence of their involvement in the program theory elaboration process. A dichotomous response screening item (Item 70) is included at the end of the previous section, asking if they were involved in any

work to elaborate the program theory and this determined whether they were asked to respond to this section of items.

The first item of the section is open response focusing on how they were involved in any theory elaboration work. The following six items; three combinations of an open response items, followed by a scaled response item, deal with how the involvement influenced their; thinking about the program (items 72 and 73), practice (items 74 and 75), and any decisions they have made or expect to make regarding the program (item 76 and 77). The three scaled response items (items 73, 75, 77) form the *Stakeholders' Perspective of the Influence of their Involvement in the Program Theory Elaboration Process Scale*. The reliability analysis of the scale is summarised below.

The literature suggests that the involvement of stakeholders in the theory elaboration process and the process of achieving consensus regarding the program theory, may increase their awareness and understanding of the program, encourage reflection regarding their practice, lead to changes in their practice, and inform decisions regarding the program evaluated (Cummings et al., 2001; Huebner, 2000; Milne, 1993). In fact, Fetterman (2004) in his elaboration of '*empowerment evaluation*,' suggests that enabling stakeholders to pursue a sound understanding of the program, the environment and the theory of action of the program (program theory as espoused) and theory of use of the program (program theory in reality) is an initial step towards empowering them to determine what changes might be necessary, as well as insight into how these changes might be implemented.

#### STAKEHOLDERS' PERSPECTIVE OF THE INFLUENCE OF THEIR INVOLVEMENT IN THE PROGRAM THEORY ELABORATION PROCESS SCALE: SCALE ITEM ANALYSIS

Data analysis of the Stakeholders' Perspective of the *Influence of their Involvement in the Program Theory Elaboration Process Scale* began with the inclusion of three items. A summary of the reliability analysis is shown in Table 5.25. Cronbach's Alpha of the three items was  $\alpha = 0.75$ . Inter-item

correlations ( $r_{ii'}$ ) for the remaining three scale items range from 0.30 to 0.71. All final scale items have an  $r_{ii'} \geq 0.3$  with two other scale items. All item-total correlations are good ( $r_{it} \geq 0.3$ ). The final scale of the three items has a mean of 7.22 and a standard deviation of s.d.= 3.03.

## **STAKEHOLDER PERSPECTIVE OF USE OF EVALUATION STUDY INFORMATION SCALES**

The format of this section is different to the other sections of this interview schedule. There are a total of 13 sub-sections in this section of the interview schedule, each focusing on a particular type of use of the evaluation information. The items of these sub-sections were intended to form the *Stakeholder Perspective of Use of Evaluation Study Information Scale*. The scale focuses specifically on stakeholders' perceptions of conceptual, instrumental and strategic influence of the evaluation information regarding the program that is the subject of the evaluation and beyond the program. The items of the *Stakeholder Perspective of Use of Evaluation Information Scale* deal with a variety of ways in which the stakeholders may have put the evaluation information to use, both formally and informally.

The 52 items of the *Use of Evaluation Study Information Scale* (Table 5.26) focus specifically on stakeholders' perceptions of conceptual, instrumental and strategic influence of the evaluation information regarding the program that is the subject of the evaluation and beyond the program. They deal with a variety of ways in which the stakeholders may have put the evaluation information to use, both formally and informally. The items have been adapted from the work of Cummings (1997), who in turn adapted some of the items from the work of Vlahov (1989) and developed others from gaps he identified from relevant literature. The 13 manners of use are the focus of this section; the items relevant to each and the type of use (e.g. conceptual, instrumental or strategic) are identified below in Table 5.26.

Table 5.25: Stakeholders' Perspective of the Influence of their Involvement in the Program Theory Elaboration Process Scale. **Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Conceptual Influence of Involvement in Program Theory Conceptualisation</b>			
ITEM 73	To what extent did this focus on program theory influence your thinking about the program?	2	0.81
<b>Instrumental Influence of Involvement in Program Theory Conceptualisation</b>			
ITEM 75	To what extent did this focus on program theory influence your practice?	3	0.54
ITEM 77	To what extent did this focus on program theory influence any decisions you've made or expect to make about the program?	3	0.46
<p><b>Note:</b> <sup>1</sup>Represents the number of other scale items the item correlates with (<math>r_{ii'} \geq 0.3</math>).</p> <p><sup>2</sup>Represents the item-total correlation (corrected) for the item with the final scale.</p>			

In the instance of this scale, the term 'Use' has been adopted in accordance with the instrument from which the items have been adapted (Cummings, 1997). The items focus on *Instrumental Use*, *Conceptual Use* and *Strategic Use* of the findings. *Instrumental Use* is the direct easily recognisable use of evaluation information to develop or improve a program (e.g. direct link to decision-making) (Weiss, 1972; Caro, 1971; Leviton and Hughes, 1981; Scriven, 1967). *Conceptual Use* refers to the influence of evaluation information on stakeholder's thinking (Leviton and Hughes, 1981; Rich, 1977). *Strategic Use*, an extension of Conceptual Use, is concerned with the use of the evaluation information to support, defend or persuade a position regarding the program (Cummings, 1993; Leviton and Hughes, 1981; Shulha and Cousins, 1997). The three types of use have been delineated with more detail earlier in this work (Chapter II).

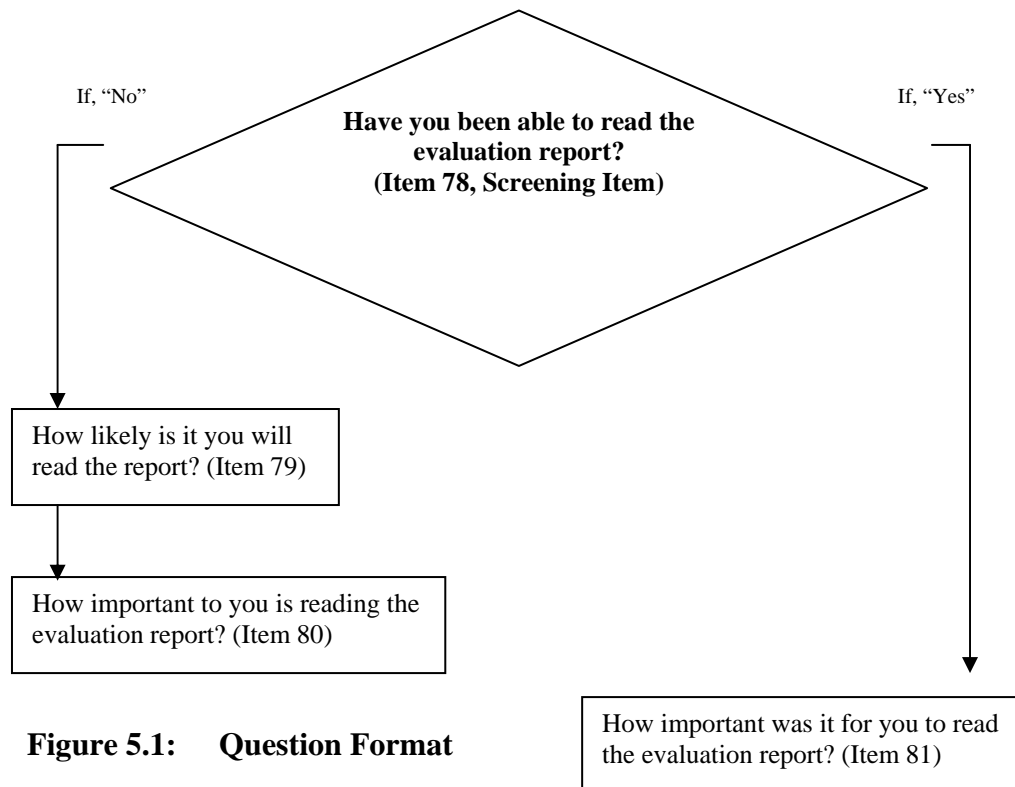
The format of this section is different to the other sections of this interview schedule. There are a total of 13 sub-sections in this section of the interview schedule, each focusing on a particular type of use of the evaluation information. The 52 items contributing to the *Stakeholders' Perspective of Use of Evaluation Study Information Scale* are composed to two sub-scales: the *Stakeholders' Perspective of the Use of Evaluation Information- Likelihood of Use Sub-scale* and the *Stakeholders' Perspective of the Use of Information-Importance of Use Sub-scale*. Each sub-section begins with a screening question regarding a particular type of evaluation information use (Figure 5.1 below includes an example of the questioning format using the first sub-section). For example, the first item of the first sub-section (Item 78) asks, "Have you been able to read the evaluation report?" The response of the interviewee then determines the follow-up items to the screening item. If the interviewee respond, "No", they are then asked the next two items of the sub-section. The second items of the sub-section deals with how likely interviewees are to put the evaluation information to use in the manner that is the focus of the sub-section (e.g. Item 79 asks "How likely is it you will read the report?"). The third item deals with how important the type of use that is the focus of the sub-section is to the interviewee (e.g. Item 80 asks "How important was it for you to read the evaluation report?"). Interviewees who have had an opportunity to use the evaluation information in the manner that is the focus of the sub-section do not respond to the second or third items of the section.

If the interviewee responds 'Yes' to the first screening item (e.g. Item 78), they are not asked the two items immediately following the screening item. They are instead asked the fourth item of the sub-section focusing on how important the type of evaluation information use that is the focus of the sub-section was to them (e.g. Item 81 asks "How important was it for you to read the evaluation report?"). Interviewees who have not had an opportunity to use the evaluation report in this manner are not asked to respond to this item.

**Table 5.26 Manner of Evaluation Information Use that are the Focus of this Section**

Manner of Use	Relevant Items	Type of Use
<ul style="list-style-type: none"> <li>▪ Reading Report</li> <li>▪ Holding informal discussions about the evaluation information with colleagues</li> </ul>	78, 79, 80, 81 82, 83, 84, 85	Conceptual Conceptual
<ul style="list-style-type: none"> <li>▪ Discussing the evaluation information at formal meetings</li> </ul>	86, 87, 88, 89	Conceptual
<ul style="list-style-type: none"> <li>▪ Undertaking further research and reading on the issues or recommendations raised in the report</li> </ul>	90, 91, 92, 93	Conceptual
<ul style="list-style-type: none"> <li>▪ Feedback on program implementation</li> <li>▪ Understanding how the program could be improved</li> </ul>	94, 95, 96, 97 98, 99, 100, 101	Conceptual Conceptual
<ul style="list-style-type: none"> <li>▪ Giving the program credibility</li> </ul>	102, 103, 104, 105	Strategic
<ul style="list-style-type: none"> <li>▪ Establishing a record of the program</li> </ul>	106, 107, 108, 109	Strategic
<ul style="list-style-type: none"> <li>▪ Influencing general policy decisions about the program</li> </ul>	110, 111, 112, 113	Instrumental
<ul style="list-style-type: none"> <li>▪ Modifying their view about the program</li> </ul>	114, 115, 116, 117	Conceptual
<ul style="list-style-type: none"> <li>▪ Making major changes to the program</li> </ul>	118, 119, 120, 121	Instrumental
<ul style="list-style-type: none"> <li>▪ Justifying the program to the community</li> </ul>	122, 123, 124, 125	Strategic
<ul style="list-style-type: none"> <li>▪ Legitimising and justifying what has been done in the program</li> </ul>	126, 127, 128, 129	Strategic

The analysis of the items required that they be coded in such a way as to represent two characteristics regarding each type of use. The first is *Likelihood of Use* and the second is *Importance of Use*. To begin, the response to the screening item (“Have you been able to...?”) was coded with either a ‘6’ (Yes) or a ‘0’ (No). The second item (“How likely is it you will...?”), asked only of those who had not yet used the evaluation information in the manner that is the focus of the sub-section, was coded with a ‘0’ for those who had read the report. For those who had not read the report, their response on a five-point scale (‘1’ representing Not at All and ‘5’ A Great Deal), was entered. The third follow-up item (“How important to you is...”) has been coded the same as the second item of the sub-section.



The coding of the stakeholder responses was undertaken in a manner slightly different to any of the items previously coded. To begin, the response to the first screening item was coded with either a ‘6’ (Yes) or a ‘0’ (No). The second item, asked only of those who had *not yet used* the evaluation information in that manner, was coded with a ‘0’ for those who *had used* the report in such a manner. For those who *had not* used the report in such a manner, their response on a five-point scale has been coded. The third follow-up item has been coded the same as the second item of the sub-section.

The fourth item of the section was asked only of those who *had used* the information from the evaluation study in this manner. A ‘0’ was given to those interviewees who *had not* used the information from the evaluation study in this manner. For those who *had used* the evaluation information their response



on a five-point scale ('1' representing Not at All and '5' A Great Deal) was entered.

The coding of the items enabled the representation of two characteristics regarding each manner of use. The first is *Likelihood of Use*. The second is *Importance of Use*. Where an interviewee chose to not respond to a sub-section because they felt that type of use was not relevant to them, they were given an '8', coded as a missing value.

To compute the *Likelihood of Use* characteristic the response code of the first item of the sub-section was added to the coded response of the second item of the section. The resulting value with regards to their *Likelihood to Use* characteristic is then somewhere on a scale from '0' to '6'. The value of '6' represents the greatest *Likelihood of Use* as the interviewee has actually used the evaluation in the manner that is the focus of the sub-section. A value of '1' to '5' represents the interviewee's perceived likelihood that they will put the evaluation information to use in this manner. For example, an interviewee who had not read the report would have been given a '0' for the first item, and a value of '1' to '5' for the second item. Their composite value for *Likelihood of Use* is then somewhere between one to five, depending on their perceived likelihood of using the evaluation information in the manner that is the focus of the sub-section. Alternatively, an interviewee who had read the report would have been given a '6' for the first item and a '0' for the second item. The resulting composite value would be a '6', the highest value for *Likelihood of Use*. The item combinations contributing to Likelihood of Use form the *Stakeholders' Perspective of the Use of Evaluation Information- Likelihood of Use Sub-scale 1 and Sub-scale 2*.

The *Importance of Use* characteristic has been calculated as a composite of the third and fourth items of the sub-section, each focusing on the importance of the particular manner of use that is the focus of the sub-section. The third item, focusing on the importance of the particular type of use for those who *have not* read the report, has been coded with a '0' for those who have read the report. A score of '1' to '5' represents the importance of putting the evaluation

information to use in the manner that is the focus of the sub-section this manner for interviewees who *have not yet done so*.

The fourth item of the sub-section, the first follow-up item for interviewees who *have used* the evaluation information in this, has been scored with a '0' for those *have not used* the evaluation information in this way. A score of '1' to '5' represents interviewees' perceived importance of putting the evaluation information to use in this manner.

To compute the *Importance of Use* characteristic the value of the third and fourth follow-up items of the sub-section have been spliced together. A value of '1' to '5' represents interviewees' perceived importance of putting the evaluation information to use in the manner that is the focus of the sub-section. The item combinations contributing to *Importance of Use* form the *Stakeholders' Perspective of the Use of Evaluation Information- Importance of Use Scale*.

The analysis of the items required that they be coded in such a way as to represent two characteristics regarding each type of use. The first is *Likelihood of Use*; the second is *Importance of Use*. The reliability analysis of the sub-scales is summarised below.

#### STAKEHOLDER PERSPECTIVE OF USE OF EVALUATION STUDY INFORMATION SCALES: SCALE ITEM ANALYSIS

Two scales have been developed to represent each of the 13 types of stakeholder use of evaluation study information that are the focus of this section; *Likelihood of Use* and *Importance of Use*. In initial attempts to undertake data analysis of the two scales the data of each was analysed as sub-scales by the type of use (conceptual/instrumental, and strategic). However, due to low Cronbach's Alpha values and low correlations ( $r_{ii'}$ ) the decision was made to analyse the 13 composite items as one scale. The analysis of each scale (*Likelihood of Use* and *Importance of Use*) is detailed separately below. Furthermore, each sub-scale has been divided into two for the purpose

of increasing the degrees of freedom for the structural equation modelling process.

Data analysis of the *Likelihood of Use Scale* regarding the use of the evaluation study information began with the inclusion of the 13 composite items (a blend of the first and second item values for each manner of use), representing the 13 manners of use listed in a summary of the reliability analysis found in Table 5.27. Cronbach's Alpha of all 26 items was  $\alpha = 0.84$ , but the deletion of the first composite (Reading Report) in the reliability analysis increased the scale Alpha to  $\alpha = 0.85$ . Inter-item correlations ( $r_{ii'}$ ) for the remaining 12 scale composite items range from -0.01 to 0.68. The final scale items have an  $r_{ii'} \geq 0.3$  with three or more other scale items. The final scale has a mean of 38.51 and a standard deviation of s.d.= 14.78.

Importance of Use. Data analysis of the *Importance of Use Scale* regarding the use of the evaluation study information began with the inclusion of the 13 composite items (a blend of the third and fourth item values for each manner of use) representing the 13 manners of use. A summary of the reliability analysis is shown in Table 5.28. Cronbach's Alpha of all 13 composite items was  $\alpha = 0.88$ . No items were deleted. Inter-item correlations ( $r_{ii'}$ ) for the final scale items range from 0.30 to 0.71. All final scale items have an  $r_{ii'} \geq 0.3$  with six other scale items or more. The final scale has a mean of 37.14 and a standard deviation of s.d.= 11.23.

#### **Stakeholder perspective of Use of evaluation study information Scales: Dividing Each Sub-scale into Two**

In the reliability analysis of the scales Item 78/79 was deleted from the *Likelihood of Use Scale* due to inter-item and item-total correlation values. This reduced the number of scale items from thirteen to twelve. For the purposes of dividing the scales into equal sub-scales, the corresponding item (Item 80/81) has been removed from the *Importance of Use Scale*. The two scales were each divided into two sub-scales, each containing six items.

**Table 5.27: Stakeholders' Perspective of the Use of Evaluation Information- Importance of Use. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS	Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Stakeholder Perception of Importance Regarding:</b>		
Reading the evaluation report (Item 80, Item 81)	8	0.50
Undertaking informal discussions about the evaluation information with colleagues (Item 84, Item 85)	9	0.64
Discussing evaluation information at a formal meeting (Item 88, Item 89)	10	0.72
Undertaking further research and reading on the issues/ recommendations raised in the evaluation report (Item 92, Item 93)	10	0.72
Using evaluation information as feedback on program implementation (Item 96, Item 97)	8	0.63
Using the evaluation information to understand how the program could be improved (Item 100, Item 101)	10	0.63
Using the evaluation information to give the program credibility (Item 104, Item 105)	4	0.38
Using the evaluation information to establish a record of the program (Item 108, Item 109)	11	0.65
Using the evaluation information to influence general policy decisions about the program (Item 112, Item 113)	10	0.72
Using the evaluation information to modify their view about the program (Item 116, Item 117)	6	0.37
Using the evaluation information to make major changes to the program (Item 120, Item 121)	7	0.43
Using the evaluation information to justify the program to the community (Item 124, Item 125)	4	0.36
Using the evaluation information to legitimating and justify what has been done in the program (Item 128, Item 129)	8	0.61
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.		

The *Likelihood of Use Scale* was divided in consideration of each item's item-total correlation value in the scale comprised of all items. The items were divided in pairs beginning with the item with the lowest value. The first sub-scale of *Likelihood of Use (L1)* is comprised of the values of item combinations 82/83, 86/87, 94/95, 102/103, 110/111, 122/123. In a reliability analysis of the sub-scale Cronbach's Alpha ( $\alpha$ ) was found to be 0.75. All items correlated with two or more items ( $r_{ii'} > 0.3$ ) with the exception of Item 122/123, which correlated with only one other item ( $r_{ii'} > 0.3$ ). Item total

correlations for all items were good ( $r_{it} > 0.3$ ), again with the exception of Item 122/123 ( $r_{it} = 0.28$ ).

**Table 5.28: Stakeholders' Perspective of the Use of Evaluation Information- Likelihood of Use. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS	Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Stakeholder Perception of Likelihood they will:</b>		
Read the evaluation report (Item 78, Item 79) Item deleted due to low $r_{ii'}$ values. When item deleted $\alpha = 0.85$ .	0	
Undertake informal discussions about the evaluation information with colleagues (Item 82, Item 83)	6	0.50
Discuss the evaluation information at a formal meeting (Item 86, Item 87)	6	0.59
Undertake further research and reading on the issues/ recommendations raised in the evaluation report (Item 90, Item 91)	7	0.53
Use the evaluation information as feedback on program implementation (Item 94, Item 95)	10	0.78
Use the evaluation information to understand how the program could be improved (Item 98, Item 99)	9	0.62
Use the evaluation information to give the program credibility (Item 102, Item 103)	5	0.45
Use the evaluation information to establish a record of the program (Item 106, Item 107)	8	0.62
Use the evaluation information to influence general policy decisions about the program (Item 110, 111)	6	0.54
Use the evaluation information to modify their view about the program (Item 114, Item 115)	3	0.40
Use the evaluation information to make major changes to the program (Item 118, Item 119)	5	0.38
Use the evaluation information to justify the program to the community (Item 122, Item 123)	3	0.36
Use the evaluation information to legitimating and justify what has been done in the program (Item 126, Item 127)	4	0.52
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.		

The second sub-scale of *Likelihood of Use (L2)* is comprised of the values of item combinations 90/91, 98/99, 106/107, 114/115, 118/119, 126/127. In a reliability analysis of the sub-scale Cronbach's Alpha ( $\alpha$ ) was found to be 0.72. All items correlated with two or more items ( $r_{ii'} > 0.3$ ) with the exception

of Item 126/127, which correlated with only one other item  $r_{ii'} > 0.3$ ). All item total correlations were good ( $r_{it} > 0.3$ ). A summary of the reliability analysis of both *Likelihood of Use* sub-scales is shown in Table 5.29.

The first sub-scale of *Importance of Use (I1)* is comprised of values 84/85, 88/89, 96/97, 104/105, 112/113, and 124/125. In a reliability analysis of the sub-scale Cronbach's Alpha ( $\alpha$ ) was found to be 0.77. All items correlated with two or more items ( $r_{ii'} > 0.3$ ). All item total correlations were good ( $r_{it} > 0.29$ ).

The second sub-scale of *Importance of Use (I2)* is comprised of values 92/93, 100/101, 108/109, 116/117, 120/121, and 128/129. In a reliability analysis of the sub-scale Cronbach's Alpha ( $\alpha$ ) was found to be 0.77. All items correlated with two or more items ( $r_{ii'} > 0.3$ ). All item total correlations were good ( $r_{it} > 0.29$ ). A summary of the reliability analysis of both *Importance of Use Sub-scales* is shown in Table 5.30.

## SUMMARY

The section above has detailed the items of the five sections of the Stakeholder Interview Schedule II. The schedule items, and the nine scales embedded, focus on stakeholders' perceptions regarding the evaluation study, the evaluators, their involvement in the evaluation study, the influence of any program theory or logic elaborated in the study or detailed in the final evaluation report, and the use of the evaluation report and its information. The schedule contains a mix of open response and closed response items. Most items requiring a response to a five-point scale address a facet of a factor for which a scale has been developed. The analysis of the scales is detailed in Chapter V.

The nine scales detailed above represent the stakeholder characteristics included in the Stakeholder Interview Schedule II expected to have influence on the use of the information of the evaluations included in this study. Interview data was used for analysis. With regards to each scale, a few items

with weak inter-item correlations ( $r_{ii'} \leq 0.03$ ) and/or weak item-total correlations ( $r_{it} \leq 0.03$ ) have been deleted. All study scales have been verified (Cronbach's Alpha  $\geq 0.7$ ). Two scales, the *Stakeholder Perceived Characteristics of Commitment to Evaluation Study- Post Sub-scale* (Sub-scale 1 of Stakeholder Perspective of Evaluation Study Characteristics) and the *Stakeholders' Perspective of the Evaluation Study Involvement Scale* have been included as scales in the Stakeholder Interview Schedule I section, therefore analysis of these scales has been undertaken in the section of this chapter dealing with the scales of that interview schedule. Data for both of these scales has been collected both pre- and post-evaluation study conduct, with a view to comparing and contrasting the influence of the evaluation process on the perspectives of stakeholders with regards to the characteristics that are the focus of these scales, and on the use of the information from the evaluation studies. The analysis of all scales helped confirm what was anticipated in terms of scale and sub-scale content and focus.

## EVALUATOR INTERVIEW SCHEDULE II

The characteristics of the evaluation as perceived by the evaluator are the focus of the Evaluator Interview Schedule II. Information regarding the characteristics of the evaluations that are the focus of this study was collected from the stakeholders of each evaluation, and from the evaluators responsible for undertaking the evaluations, before the conduct of the study. The purpose of this interview schedule is to gain insight regarding the evaluation from the perspective of those responsible for undertaking the evaluation, after it had been conducted.

Due to the low number of evaluators interviewed (seven), where there is a missing value, an imputed value has been included. In each case this is the mean of the other respondents' responses to that scale item. This has been done so as to maintain the full number of cases for data analysis. In fact, this was done for two cases and with regards to only one item. In both instances, the evaluator joined the evaluation team after the conduct of the data collection and felt uncomfortable answering Item 4a regarding the extent to which the evaluation team were given sufficient latitude to exercise sound professional judgment in undertaking the evaluation.

**Table 5.29: Stakeholders' Perspective of the Use of Evaluation Information- Likelihood of Use Sub-scales. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS	Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Likelihood of Use Sub-scale 1</b>		
Undertake informal discussions about the evaluation information with colleagues (Item 82, Item 83)	3	0.53
Discuss the evaluation information at a formal meeting (Item 86, Item 87)	3	0.62
Use the evaluation information as feedback on program implementation (Item 94, Item 95)	4	0.74
Use the evaluation information to give the program credibility (Item 102, Item 103)	2	0.34
Use the evaluation information to influence general policy decisions about the program (Item 110, Item 111)	3	0.48
Use the evaluation information to justify the program to the community (Item 122, Item 123)	1	0.28
<b>Likelihood of Use Sub-scale 1</b>		
Undertake further research and reading on the issues/ recommendations raised in the evaluation report (Item 90, Item 91)	2	0.34
Use the evaluation information to understand how the program could be improved (Item 98, Item 99)	4	0.60
Use the evaluation information to establish a record of the program (Item 106, Item 107)	4	0.60
Use the evaluation information to modify their view about the program (Item 114, Item 115)	2	0.41
Use the evaluation information to make major changes to the program (Item 118, Item 119)	3	0.48
Use the evaluation information to legitimating and justify what has been done in the program (Item 126, Item 127)	1	0.40
<b>Note:</b> <sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale.		



**Table 5.30: Stakeholders' Perspective of the Use of Evaluation Information- Importance of Use Sub-scales. Item Analysis Results.**

SCALE ITEMS GROUPED BY AREA OF FOCUS	Number of Items $r_{it} \geq 0.3^1$	$r_{it}^2$
<b>Importance of Use Sub-scale 1</b>		
Undertaking informal discussions about the evaluation information with colleagues (Item 84, Item 85)	3	0.55
Discussing evaluation information at a formal meeting (Item 88, Item 89)	3	0.68
Using evaluation information as feedback on program implementation (Item 96, Item 97)	4	0.59
Using the evaluation information to give the program credibility (Item 104, Item 105)	2	0.29
Using the evaluation information to influence general policy decisions about the program (Item 112, Item 113)	4	0.69
Using the evaluation information to justify the program to the community (Item 124, Item 125)	2	0.34
<b>Importance of Use Sub-scale 2</b>		
Undertaking further research and reading on the issues/ recommendations raised in the evaluation report (Item 92, Item 93)	5	0.63
Using the evaluation information to understand how the program could be improved (Item 100, Item 101)	5	0.58
Using the evaluation information to establish a record of the program (Item 108, Item 109)	4	0.62
Using the evaluation information to modify their view about the program (Item 116, Item 117)	2	0.30
Using the evaluation information to make major changes to the program (Item 120, Item 121)	4	0.50
Using the evaluation information to legitimating and justify what has been done in the program (Item 128, Item 129)	3	0.48

A majority of the items in the interview schedule are in open response format (items 1, 2b, 3b, 4b, 5b, 6, 7 and 8). The information collected in response to these items is not appropriate for contribution to a scale. However, it is reported later in the dissertation to add insight to the quantitative information.

The Evaluator Interview Schedule II begins with a series of open response items (Item 1 a-g) focusing on the collection of information regarding the evaluator's involvement with the evaluation study and their perception of the purpose, methods and information collected in the study. These items are the

same as those included in Item 6 of the Evaluator Interview Schedule I, with some changes to tense wording. The inclusion of the same set of items in the two interview schedules enables the comparison and contrast of evaluator responses given at the beginning of the study and at the end of the study, with a view to determining any differences.

The next four items of the interview schedule (items 2a, 3a, 4a and 5a) are scaled response format focused on gaining insight into evaluator perception regarding the evaluation process in terms of the extent to which they felt their skills were appropriate (Item 2), the evaluation team was competent to undertake the evaluation study (Item 3), the evaluation team was given sufficient latitude to exercise sound professional judgment in undertaking the evaluation (Item 4), and the extent to which they felt the evaluation to have been a negative or positive experience (Item 5). Items 2a, 3a, 4a and 5a require a response on a five-point scale. The terminology attached to each scale varies in accordance with the nature of the question. For instance, Item 2a asks evaluators, “To what extent do you feel your skills were appropriate to this evaluation?” on a five-point scale, ‘1’ representing ‘Not Appropriate’ and ‘5’ representing ‘Very Appropriate’. Similarly, the end-points of the response scales of items 3a, 4a and 5a are ‘Not Competent/Very Competent’, ‘No Latitude/A Great Deal of Latitude’, and ‘Very Positive Experience/Very Negative Experience’, respectively. The four items (Items 2, 3, 4, and 5) have been included to form the *Evaluator Perceived Characteristics of the Evaluation Scale*. Each of these four items has a sub-item of open response format included to gain further insight into why evaluators’ have responded as they have to the scaled response format item.

The final three items of the interview ask the evaluators to consider the ways the primary stakeholders of the evaluation may have been influenced by: the evaluation process (Item 6); the findings (Item 7); and the ways they feel the evaluation information will be put to use (Item 8). Item 9 is a final item for evaluators to respond with any further comments they felt necessary. These final four questions are all open response items.

## EVALUATOR PERCEPTION OF THE EVALUATION CHARACTERISTICS SCALE

Only four quantitative items have been included in the Evaluator Interview Schedule II to gain insight into evaluators' perspective of the evaluation study characteristics. These four items were intended to form the *Evaluator Perception of the Evaluation Characteristics Scale*. The reliability analysis of the scale is summarised below.

### EVALUATOR PERCEPTION OF THE EVALUATION CHARACTERISTICS: SCALE ITEM ANALYSIS

The four items focusing on this area were conducted to form the *Evaluator Perceived Characteristics of the Evaluation Scale*. The initial Cronbach's Alpha value of all four items is  $\alpha = 0.14$ , only rising to  $\alpha = 0.54$  with the deletion of Item 5a. Unfortunately, the low Cronbach's Alpha value of the scale and weak inter-item correlations (only items 3a and 4a correlated with each other with  $r_{ij} \geq 0.3$ ) indicate that the items do not form a sound scale. However, the data collected in response to these items is reported later on an item-by-item basis to add insight to the findings of this investigation. A summary of the reliability analysis undertaken with this scale is presented in Table 5.31.

## SUMMARY

The section above has summarised details of the nine items included in the Evaluator Interview Schedule II. The items are a mix of open response and scaled response format, with four of the items contributing to one scale. They focus on evaluators' thinking regarding the evaluation process, the influence of the evaluation process and findings on stakeholders, and ways they feel the evaluation information will be put to use.

**Table 5.31 Evaluator Perceived Characteristics of the Evaluation Scale**

SCALE ITEMS GROUPED BY AREA OF FOCUS		Number of Items $r_{ii'} \geq 0.3^1$	$r_{it}^2$
<b>Evaluator Perception of Fit with Evaluation Needs</b>			
ITEM 2a	To what extent do you feel your skills were appropriate to this evaluation?	0	0.10
ITEM 3a	To what extent do you feel the evaluation team was competent to undertake the evaluation study?	1	0.43
<b>Evaluator Perception of Evaluation Environment</b>			
ITEM 4a <sup>3</sup>	To what extent do you feel the evaluation team was given sufficient latitude to exercise sound professional judgment in undertaking the evaluation?	1	0.07
ITEM 5a	Thinking about the evaluation of _____ program, to what extent do you consider the evaluation to have been a negative or positive professional experience?	0	-0.16
<b>Note:</b>		<sup>1</sup> Represents the number of other scale items the item correlates with ( $r_{ii'} \geq 0.3$ ). <sup>2</sup> Represents the item-total correlation (corrected) for the item with the final scale. <sup>3</sup> Missing responses to this item for two cases have been replaced by the mean of all other responses to this item.	

Interview data were used for analysis of the Evaluator Perception of the Evaluation. Missing data (two responses) for Item 4a was replaced with the mean of the responses of all of the other evaluators to that item with a view to maintaining the number of cases (seven) included in the analysis. The low Cronbach's Alpha value of the scale and weak inter-item correlations (only items 3a and 4a correlated with each other  $r_{ii'} \geq 0.3$ ) indicate the items do not form a sound scale. Therefore, the responses to the items of this scale will be included later in this dissertation on an item-by-item basis to add further insight to other qualitative and quantitative data gathered in this study.

## CONCLUSION

The Organisational Characteristics Interview Schedule focus of this chapter has been on the six instruments of the study. These are the Stakeholder Interview Schedule I, Evaluator Interview Schedule I, Process Use Interview Schedule, Stakeholder Interview Schedule II, and the Evaluator Interview Schedule II. The logic underpinning the development of each interview schedule and its items, both qualitative and quantitative, is considered. The analysis process of the scales of each of the interview schedules from which the observed variables of the Core Model and the further elaborated model of the study were drawn, is described and changes made to the scales as a result is detailed. All scales have been verified (Cronbach's Alpha  $\geq 0.7$ ). Items with weak inter-item correlations ( $r_{ij} \leq 0.03$ ) and/or weak item-total correlations ( $r_{it} \leq 0.03$ ) have been deleted from scales, with consideration given to the contribution of the item to the range of concepts that are of focus in each scale.

Chapter VII is concerned with the development of the structural equation model of the study. The latent variables of the Intended Elaborated Model are *Program Theory* (Causative or Implementation), *Commitment to the Evaluation Study* (Pre), *Commitment to the Evaluation Study* (Post), *Evaluation Study Characteristics*, *Process Use*, *Organisational Environment Characteristics*, *Stakeholder Characteristics*, *Evaluator Characteristics*, *Expected Study Involvement* and *Study Involvement* (Post). All variables, with the exception of the *Program Theory* variable, are taken from the scales of the interview schedules. The *Program Theory* variable is based on classifications of program theory used in the evaluation case studies described in the next chapter.

## **CHAPTER VI**

### **THE EVALUATION STUDIES**

This chapter contains three main sections, each dedicated to one of the programs and its evaluation study which were the ‘cases’ examined in this research. Each program is described together with its primary stakeholders. The program theories guiding and underpinning the programs, including both their implementation and causative theories, are then considered. Next, the evaluation study of the program is focused on, and the aims or intentions of the study are detailed. Finally, each evaluation’s use of the program’s theory is classified using the Classification Matrix, developed in Chapter III of this work. A table representing the classification of the three evaluation studies is presented at the end of the chapter.

The program stakeholders are acknowledged in the description of each program. Scriven (1991) defines a stakeholder as someone with a “...substantial ego, credibility, power, futures, or other capital invested in the program, and thus can be held to be to some degree at risk with it” (p. 334). Stakeholders are integral to this study for four main reasons. First, stakeholders are often in a position to influence a program’s implementation in some way, either directly or indirectly. For instance, the program manager is integral to the delivery of the program and typically in a position to affect a program. While the sponsors of a program are less directly involved in program delivery, the conditions they attach to the granting of funds may have a substantial impact on how the program is run. Second, program stakeholders may have an impact on how the evaluation of the program is undertaken. Stakeholders directly involved with the program might be involved in the evaluation process, either as part of a reference group or as an information source, and therefore clearly in a position to influence the evaluation. Stakeholders who are less directly involved, for example program sponsors, might influence the evaluation by requiring certain information to be made available to them regarding the manner in which the funds have been used (i.e.

accountability). Furthermore, the use of the stakeholders' theory of the program in the evaluation process is a variable considered in this work. Third, program stakeholders are the users of the evaluation's information, and evaluation information use is the primary dependent variable of this study. Finally, stakeholder interviews are the primary source of information for the study, forming the backbone of this work.

The theory guiding each program has been detailed below. Both program implementation theory and causative theory have been elaborated. Definition and the detailed consideration of both program causative theory and program implementation theory, the two major types of theory use upon which the *Classification Matrix* is based, have been provided in Chapter II of this work. In the simplest terms, a program's causative theory identifies the causal links between a program's application and its intended effects. Alternatively, implementation theory considers how the program is conducted, and addresses the variables affecting the extent to which the program is implemented as planned (Scheirer, 1987; Weiss, 1997b).

It is important to note here that the classification of the 21 studies undertaken earlier and presented in the elaboration of the Classification Matrix was based solely upon the journal articles reporting those evaluations. Therefore, the information upon which the classifications rest is actually a summary of the evaluation process and likely, in many cases, to have been synthesized and summarised for publication. With regards to the Pre-Apprenticeship, YOHFest and Making Consistent Judgments (MCJ) program evaluations, examined in this research, the program theory-use classifications presented below are based on a number of information sources including interim evaluation reports, documents produced in meetings (e.g. theory elaboration with program stakeholders), formal interviews, informal conversations with evaluators and program stakeholders, and the final evaluation reports. Having the additional information sources to classify the evaluations has made the classification process a little more complex, yet more informed.

## YOUTH ON HEALTH FESTIVAL

YOHFest is a festival in which students from high schools throughout the State of Western Australia perform plays they have written and choreographed, perform a dance, or enter an art piece that they have created. The primary aims of the program are to:

- i. Increase the knowledge and skills and develop positive attitudes among young people regarding health issues;
- ii. Increase their life skills; and
- iii. Enhance the coping skills and resilience of young people which will foster their successful transition from adolescence to adulthood.

The festival has been developed around the logic that peer education is an effective method of addressing youth health issues. This logic was derived from relevant social science theory regarding peer education and stakeholder theory. The primary stakeholder of the festival, the person who initiated and instigated the development of the concept, has a nursing background and a lengthy history of working in the health promotions area. She is aware of the relevant research in the area and also has strong practical experience upon which she has based her thinking in developing the festival.

The logic underpinning YOHFest (YOHFest, 2004) is that in the transition from adolescence to adulthood, the peer group offers the security of group membership. During this time of transition, the influence of the peer group on youth values and behavioural standards is strong. Capitalising on this influence, peer education programs use peers to deliver reliable information in an environment where open discussion is valued. One program utilising the peer education strategy considered in the development of the YOHFest program is “Promoting Adolescent Sexual Health (PASH)”, delivered through Family Planning of Western Australia ([www.fpwa-health.org.au/pash.htm](http://www.fpwa-health.org.au/pash.htm), 25/11/2004). The logic of YOHFest rests on the belief that peer education programs are more influential than professionals in the delivery of educational and counselling services (YOHFest, 2004).



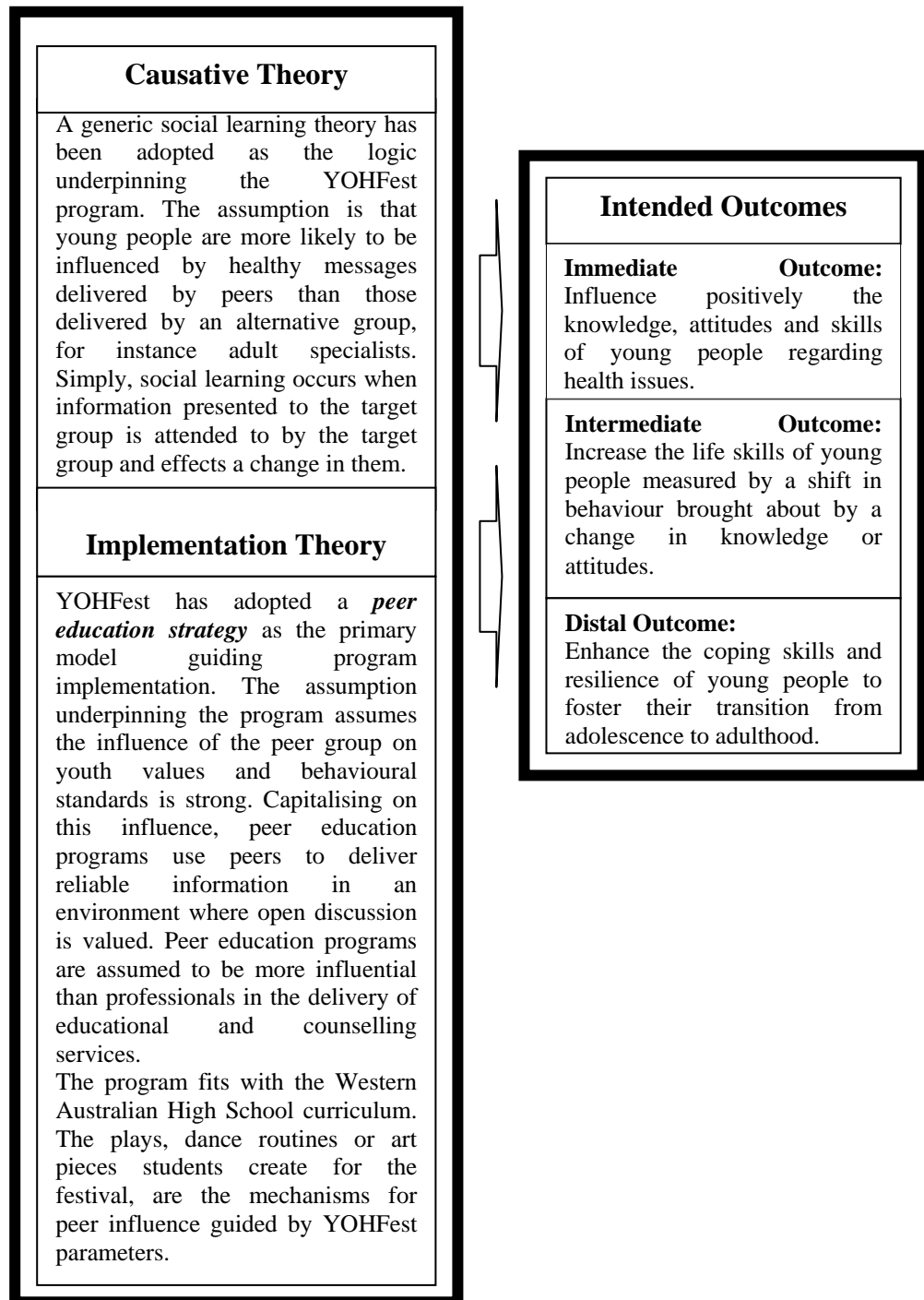
Two themes were addressed in the festival entries for 2004. The primary theme for the festival was “You + Me + Acceptance = Equality”. The second theme was “Smarter than Smoking”. The plays were to be between 15 and 20 minutes in length, the dance routines approximately eight minutes in length and the art piece one metre square.

### **YOHFest Primary Stakeholders**

Each member of the YOHFest stakeholder group became involved in the festival for a reason slightly different to the others. For instance, Healthway and ALCOA provided sponsorship money. The Lions Club provided resources and oversight of the young people, managing the group through their involvement in the Leo Club. The Department of Education and Training in Western Australia was concerned that each student should reach their learning and skill potential and contribute to society. The festival provides an opportunity for high school students to do so, in keeping with that aim. The teachers who worked with the students in the development of their presentations were responsible for the education of the students in line with the Western Australian Curriculum Framework. The students involved themselves in the development of their festival presentation, presumably with some motivation to learn new knowledge and skills. However, they may also have had less obvious motivators such as the desire to fit in with a particular peer group, or simply to fulfil an educational requirement.

The aims, influences and theories of the various stakeholder groups have been more carefully delineated below. The consideration of the YOHFest stakeholders individually facilitates a clearer understanding of YOHFest as realised in its implementation by accounting for the influential variables. Interestingly, although each stakeholder group had a different interest in the festival, there were similarities in terms of what they hoped to achieve by sponsoring, or becoming involved, in YOHFest. Considering the place of each stakeholder in relation to the program helps to contextualise YOHFest. **Lions Club.** YOHFest has a very strong link with the Lions Club of Mandurah, Western Australia. In fact, the Healthway Arts Sponsorship award for the over

\$5000 category was made to the “Lions Club of Mandurah for its 2003 Youth on Health Fest”. As mentioned earlier, the incorporation of YOHFest in 2004 was intended to foster the separation between Lions Club of Mandurah and the Festival. Nevertheless, the roots holding them together go very deep.



**Figure 6.1: YOHFest Program Theory Representation**

The YOHFest Management Team is composed of around 40 “Leos”, the young people’s contingent of the Lions Club, linked to the Lions Club of Mandurah in Western Australia and a few adult members of the club. The structure of the Lions Clubs in Australia is such that a Lions Club oversees all Leos Clubs and the parent Club determines candidate eligibility for membership of Leos Clubs. The Lions Club of Mandurah screens young people in considering their eligibility for Leo membership. In effect, they provide a screening service which ensures that the youth involved in the YOHFest Management Team are of good character and open to accepting the values of the Lions Club. The chosen Leos then have an opportunity to develop and use their leadership skills in the management of a forum where young people, Leos included, are supported in taking on board healthy messages that will enable them to make healthier life choices or decisions.

The values of the Lions Clubs and the Leos are in keeping with the values underpinning YOHFest. In particular, the development of the Leos leadership skills is strongly supported by the Mandurah Lions Club. YOHFest offers a good opportunity for the development and establishment of leadership skills. In keeping with Lions Club International, the Lions Club motto is “We serve” and the mission of the Club is to “create and foster a spirit of understanding among all people for humanitarian needs by providing voluntary services through community involvement and international co-operation.” ([http://www.lionsclubs.org/EN/content/lions\\_mission.html](http://www.lionsclubs.org/EN/content/lions_mission.html); 11/11/04).

**South Metropolitan Health Unit.** The aim of the Department of Health is to manage a “comprehensive range of health and health-related services to all West Australians, including health promotion, health protection, diagnosis, treatment, rehabilitation, continuing care, support and palliative care...” (<http://www.health.wa.gov.au>; 25/11/2004). The Department’s Health Promotion program, in particular, is concerned with promoting lifestyle changes to avoid premature illness, disability and early death, through education of people in ways which develop and maintain environments supportive of health enhancing behaviour and reduce behaviour compromising to their health.

Resources and funds were provided to YOHFest by the Department of Health through the South Metropolitan Health Unit. The Health Promotions Unit, in particular, monitored the festival developments and has an interest in the event. The person responsible for the elaboration of the festival concept and implementation until the beginning of 2004 was employed by the South Metropolitan Health Unit in the capacity of a school nurse responsible for its health promotion. In fact, the YOHFest office was located within the premises of the Mandurah Community Health Centre. This situation may not continue as she is no longer employed by the South Metropolitan Health Unit and now works full-time as part of the YOHFest Management Team.

The interest of the Department of Health (DoH) in YOHFest is primarily through its promotion of healthy messages to youth. The DoH also has an interest in improving the health and well-being of aboriginal people and people living in remote areas of the State. YOHFest addresses this aim as it endeavours to involve such groups in the Festival.

Although YOHFest has very strong links with DoH, the Department has little influence on the delivery or management of the festival. However, it does monitor it closely. If YOHFest was not in line with DoH aims, or implemented in a manner it found unacceptable, they would not support the festival. Furthermore, although YOHFest has been developed and elaborated by a DoH employee, and the YOHFest office is located on DoH premises, it operates as its own entity. Had the festival been at odds with DoH, this facilitative environment would not have been made available.

**Department of Education and Training.** The purpose of the Department of Education and Training (DET) is to ensure that all school students develop the knowledge, skills and confidence needed to achieve their individual potential and contribute to society. Although DET has not officially flagged YOHFest as a project to be considered by schools or teachers, it is aware that school students enter the festival.

One drama teacher with a history of considerable involvement in YOHFest has developed a workshop for presentation to drama teachers focusing on how YOHFest can be incorporated into the curriculum. Some teachers have attended her workshop. This teacher and the YOHFest Management Team continue to lobby for the festival to be included as part of the high school curriculum. This would encourage more teachers to take on the development of entries for the festival and increase the numbers presenting at the festival. Although the DET has allowed this work to continue, they have not officially endorsed YOHFest. One reason for this is likely to be the number of similar programs that would also demand a level of support.

DET is the gatekeeper. It has allowed YOHFest as a project for students because the aims of the festival are compatible with the over-arching aim of DET. Furthermore, it is compatible with the Western Australian Curriculum Framework. The Western Australian Curriculum Framework embodies the broad-based causative and implementation theories underpinning the education of students in Western Australia. YOHFest represents one avenue by which the framework for student education may be implemented. The primary causative theory underpinning YOHFest, the use of a peer education strategy to deliver healthy messages, is compatible with that of DET. The broader YOHFest implementation theory regarding students' development and presentation of a drama, dance, or art piece in a festival setting is also compatible with the broader education strategy adopted by DET. However, DET does not have a significant influence on the manner in which the festival is implemented. However, it does have an influence in terms of enabling a context in which the festival is able to proceed without significant interference. It is understood that if DET felt student involvement in the festival was at odds with its overall policies and objectives, then youth groups from the schools would not present. Youth groups would have to be involved through other agencies or organisations.

**Healthway.** Healthway, a primary sponsor of YOHFest, was established in 1991 under the *Tobacco Control Act 1990* of Western Australia as a statutory body. The objectives of Healthway are to fund activities promoting health,

particularly of young people, provide research grants to organisations undertaking health promotion programs and research, and offer sponsorship support for sports, arts and racing activities which encourage healthy lifestyles and provide opportunities to promote healthy messages.

The objectives of YOHFest and the manner in which it is undertaken are very compatible with those of Healthway. With regards to arts projects in particular, Healthway sponsors arts activities which provide opportunities for health promotion and access to a wide range of audiences. The main objective of Healthway is “...the promotion of health with a particular focus on young people, indigenous groups and rural and remote communities in Western Australia” (<http://www1.healthway.wa.gov.au>; 25/11/2004). In 2004 Healthway chose YOHFest as the winner for the Arts Sponsorship over \$5000 Award for its 2003 festival.

In conversation with a representative of Healthway it was revealed that the organisation is particularly happy with the YOHFest strategy because not only does it promote healthy messages, it also aims to involve young people from all schools across Western Australia, including indigenous groups and rural and remote community groups. Furthermore, the healthy message is actually a focus of the YOHFest event. This is rare. Most often Healthway-sponsored programs and projects promote healthy messages through the event or program, but are not integral to its undertaking, in contrast to YOHFest.

Although Healthway provides sponsorship to YOHFest they do not have a hands-on role in the delivery of the festival. The use of a peer education strategy to manage an arts event based on a healthy message theme is particularly compatible with the aims and objectives of Healthway. However, it does have an influence on the implementation of the festival in less obvious ways. For instance, Healthway's concern regarding the health of indigenous groups and rural and remote community groups has found its way into the implementation of YOHFest. It is unlikely that the inclusion of these groups in the festival would have occurred without the Healthway incentive. The provision of additional Healthway sponsorship money to projects or programs

endeavouring to promote health to these groups in particular, may have furthered YOHFest development in this area.

**ALCOA.** A large number of ALCOA employees and their families reside in Mandurah as ALCOA has four production and mining sites located close to and around this community. Although ALCOA has for many years extensively supported environmental, conservation and land care projects, it has only recently begun to place a stronger focus on the development of community sponsorship partnerships and programs. The company pledges to form and undertake a range of sponsorships to strengthen communities under the recently initiated sponsorship banner of “Partnering Stronger Communities” (<http://www.alcoa.com/australia>; 25/11/2004).

Based in Mandurah, YOHFest falls within the range of projects and programs ALCOA has chosen to sponsor. ALCOA has an interest in the aims and strategies of the festival in terms of understanding what it is they are funding, and assessing the appropriateness of the festival to the community-building programs they sponsor. For ALCOA, YOHFest represents a tool for the implementation of their stronger community-building initiative and a way for the company to contribute to the community. Broadly, the promotion and education of healthy messages to the youth of the community and the development of leadership skills among Mandurah youth, works towards the growth of a stronger and healthier community.

## **YOHFEST PROGRAM THEORY**

The intent of this section is to describe the causative theory and implementation theory of the YOHFest program. The causative theory, or the theory underpinning the program, is considered first, followed by a focus on implementation theory, or the theory linking the activities of the program.

**Causative Theory.** The primary causative theory underpinning the program is based on the assumption that youth are more likely to attend to and trust information from peers. The festival has been developed around the logic that

peer education is an effective method of addressing youth health issues. This logic was derived from relevant social science theory and stakeholder theory. In the transition from adolescence to adulthood, the peer group offers the security of group membership. Research, such as that by Turner and Shepherd (1999), indicates that the influence of the peer group on youth values and behavioural standards is strong. Capitalising on this influence, peer education programs use peers to deliver reliable information in an environment where open discussion is valued. Furthermore, the same body of research suggests that peer education programs are more influential than professionals in the delivery of educational and counselling services.

The person originally responsible for the conceptualisation and development of the YOHFest concept has a nursing background and a lengthy history of working in the area of health promotion. She is aware of the relevant research in the area and also has strong practical experience, as a school nurse responsible for health promotion, upon which she has based her thinking in developing the festival.

**Implementation Theory.** YOHFest is managed and run by youth. Students from high schools throughout the State of Western Australia perform plays they have written and choreographed, perform a dance, or enter an art piece they have created. The use of art as a medium for the festival enables the young people involved to engage with the healthy message and present their interpretation at the festival competition. In some instances, they also present to their own school audience and members of their local community.

The willingness of drama teachers of school to work on the development of a YOHFest item within the parameters of the curriculum is imperative. School involvement in the festival is also negotiated, however, through the willingness of the schools, teachers and students to participate, and the availability of school resources.

The YOHFest Management Team (MT) is composed of Leos, the young people's contingent of the Lions Club. A limited number of adults are also part



of the MT. Membership in Leos is open to young men and women of good character who are deemed eligible by the Leo Club committee of the sponsoring Lions Club. The Leo group responsible for the management of YOHFest, is linked to the Lions Club of Mandurah, Western Australia. The MT liases with the schools, the community and sponsors in the management of the festival and development and the annual event. They give talks at schools and other community information forums in order to increase participation and gain support.

In addition to undertaking the day-to-day administrative duties associated with the festival throughout the year, the MT also works as a team during the festival to ensure it is presented professionally in a friendly and supportive environment. The MT attends at least two weekend camps each year to, among other things, develop group cohesiveness, encourage the development and practice of leadership skills, and develop positive attitudes for healthy choices.

Through attachment to the Lions Club of Mandurah, YOHFest is also able to make use of the Lions' resources and benefits, such as the Lions Club network and insurance, in the running of the festival. In 2004, however, YOHFest has taken steps to become incorporated, and stand as a legal entity in its own right. Although to date this has had little impact on festival implementation, changes may unfold with time, such as the resources and benefits made available through the Lions Club of Mandurah.

The implementation of YOHFest also rests on the availability of sponsorship money and VIPs to act as patrons of the festival. High profile patrons of good standing in the community support YOHFest publicly. Their public commitment to the festival raises awareness and the perceived credibility of the event.

## **YOHFEST THEORY USE CLASSIFICATION**

The purpose of this section is to describe the classification of the program theory use in the evaluation of the YOHFest program. The classification is undertaken in accordance with the Classification Matrix presented in Chapter III (Table 3.1).

A process of program theory elaboration was undertaken early in the evaluation process by the YOHFest program evaluation stakeholder group led by the primary evaluator. The group then negotiated the evaluation information focus and instrument development. The classification of the YOHFest program theory outlined below, rests on a review of these instruments and consideration of information from interviews with the key stakeholders of the program and its evaluation.

The festival had been developed around the logic that peer education is an effective method of addressing youth health issues. Because there is strong use of causative theory to drive the program, the use of causative theory in the undertaking of the evaluation was important to the stakeholders. In fact, the questionnaire developed by the evaluators in consultation with the key stakeholders, primarily focuses on obtaining information from YOHFest participants, regarding shifts in their knowledge and attitudes linked to the festival themes. A review of the data collection instruments indicates a strong focus on attitudinal shifts both from YOHFest participants and Leo Management Team members. In other words, a focus on the causal links between the implementation of the program and expected program outcomes or effects. A key YOHFest stakeholder said in an interview that this was the sort of information she required to justify funding.

The causative theory underpinning the program was derived from relevant social science theory and stakeholder theory. In a theory elaboration process undertaken early in the evaluation conceptualisation stage, this theory was detailed and organised by the evaluators for stakeholder feedback. The development of the data collection instruments was then based upon this

theory. Social Science Theory, Stakeholder Theory, and Evaluator Theory have all been used in a *Substantial* manner in the development of the causative theory guiding the evaluation.

In contrast to the situation for causative theory, a review of the evaluation surveys indicates that little information was collected regarding the implementation of the program. Again, in interviews with key program stakeholders, it was stated that they were happy with program implementation and were working to increase the number of schools and the geographical areas from which they came. Specifically, they were interested in gaining information from the evaluation to encourage sponsorship, with a view to growing the program. They needed to show that the attitudinal shifts expected to occur in participants through involvement with the festival, were occurring. All three types of theory have been given a *None or Negligible* rating with regards to Implementation Theory.

With regards to context, little or no specific contextualisation of survey items has been undertaken. The items are such that contextualisation is not necessary. The items primarily address common youth health issues such as alcohol use, drug use, smoking and sex. The contribution of Social Science Theory, Stakeholder Theory and Evaluator Theory have all been guided by an understanding that the evaluation target group are youth. In this instance, a classification of C1 has been given with regards to all three types of theory sources, as consideration was given to the target population's age. The theory-use classification of YOHFest is summarised in Table 6.1.

#### THE PRE-APPRENTICESHIP TRAINING PROGRAM IN WESTERN AUSTRALIA

Pre-Apprenticeships are one of a number of training opportunities available to young people. Pre-Apprenticeships were originally intended to be entry-level courses leading into trade-specific employment and/or further training upon completion (NCVER: 2003). Unlike the traditional apprenticeship where the student is indentured to an employer, Pre-Apprenticeships are primarily

institution-based and include a work experience component. The three sectors of program provision include TAFE, schools and private providers.

**Table 6.1: Theory-Use Classification of the YOHFest Evaluation**

	Program Implementation Theory <i>Program Theory Source</i>									Program Causative Theory <i>Program Theory Source</i>								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
Program Evaluation Study	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial
YOHFest Evaluation	NM			NM			NM					C 1			C 1			C 1
<p><b>Note:</b> NM in a box indicates the theory origin was not mentioned in the information sources included in the investigation.  C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based.  The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).</p>																		

In 1999 Workplace Skills Management undertook a review of Pre-Apprenticeships in Western Australia. They found:

- i. Some students were significantly over-trained;
- ii. Articulation from some courses was poor due to limited or non-existent industry demand;
- iii. Unsatisfactory student outcomes; and,
- iv. Industry reluctance to accept Certificate II level qualification due to the associated award cost implications.

These findings lead to the development and implementation in 2001 of a number of recommendations regarding Pre-Apprenticeships. These included:

- i. The reduction of structured competency hours to no more than 600;
- ii. The inclusion of the Pre-Apprenticeships in the Department of Training's planning processes as part of the State Training Profile;

- iii. The reduction of Pre-Apprenticeships to Level 1 certificate; and,
- iv. The inclusion of generic competencies.

A statistical analysis of TAFE enrolments in 2001 and 2002, undertaken in August 2003 (Department of Education and Training; 2003) found that although the changes to the Pre-Apprenticeship Program had addressed the problem of over-training, a significant increase in course enrolments had occurred that was at odds with the static number of apprenticeships. The report that resulted from this analysis recommended that an evaluation be undertaken to assess the Pre-Apprenticeship outcomes, including employment pathways and value for money.

Coincidentally, recent public recognition of the shortage of skilled tradesmen has encouraged the current socio-political environment to lean towards favouring Pre-Apprenticeship programs in terms of improving the supply of trades' people. The shortage of trades people has been brought about by: changes in workplace practice, where the move towards sub-contracting has meant fewer operators have the capacity to take on an apprentice; the lack of trades people to replace the 'baby-boomer' generation who are retiring; growth in the economy creating a demand for skilled trade people to build and maintain infrastructure capacity; and, the increasing unwillingness of young people to enter what has become known as the 'dirty' careers (Zapelli et al., 2004, September).

#### **PRE-APPRENTICESHIP PROGRAM STAKEHOLDERS**

The primary stakeholders of the program are considered individually below.

**Department of Education and Training (DET).** DET is responsible for planning Western Australia's long-term training system to meet the needs of the community, industry and regional areas. Furthermore, the aim of DET is to provide life-long training and employment opportunities to all Western Australians (<http://www.training.wa.gov.au>; 14/06/05). The programs DET dedicates funds to are those that further the implementation of the

Department's policies. The Pre-Apprenticeship Program is one program DET funds. Program funding is negotiated annually in meetings between DET and TAFE representatives. Although the individual courses are reviewed by TAFE, DET is accountable for the management of its funds. As such, DET is responsible for overall evaluation and monitoring of the Pre-Apprenticeship Program.

**TAFEWA.** TAFEWA offers a secondary study option to Western Australians as an alternative to the traditional classroom-based schooling options. TAFE offers a range of learning options based on a mix of classroom and practical placement (<http://www.tafe.wa.gov.au/HowTAFEWAWorks/philosophy>; 14/06/05). TAFEWA is the primary service provider of the Pre-Apprenticeship Program as DET funds TAFEWA colleges to deliver the courses. Not all colleges deliver courses in the same skill areas. In fact, some specialise in particular areas such as electrical or plumbing, while others offer courses covering a more diverse range.

**The State Training Board (STB).** The STB is the primary industry and community advisory body established to provide leadership and strategic advice for vocational education and training in Western Australia. The Board, established under the *Vocational Education and Training (VET) Act 1996*, provides advice to the Minister for Education and Training regarding strategies and policies aimed at meeting the learning needs of Western Australians. The STB carries the perspective of industry and community into collaboration with DET. The intention of the STB is not to duplicate the role of DET, but to augment it (<http://www.stb.wa.gov.au/about/content.htm>; 14/06/05). Regarding the Pre-Apprenticeship Program, the STB has a role in overseeing the implementation and process of the program with a view to making comment as the body responsible for providing strategic advice on vocational education and training in the State.

**Training Accreditation Council (TAC).** The TAC is responsible for accrediting industry organisations (<http://www.tac.wa.go.au/Operating.html>; 14/06/05) for training. These registered training organisations (RTO) function

as service providers for the Pre-Apprenticeship Program. RTOs are the primarily practical placement options for Pre-Apprenticeships. RTOs, in addition to providing practical placement opportunities, have links into other industry placement. The TAC is responsible for monitoring the organisations recognised as training providers, including their dealings with Pre-Apprenticeship students.

**Industry.** The RTOs are industry-based organisations responsible for the practical training of a portion of Pre-Apprenticeship students. They have a stake as service providers of the program. In addition, the goal of Pre-Apprenticeship students is to work as skilled tradespersons in an industry area. The various industry areas have a stake in the Pre-Apprenticeship Program in that they may eventually be responsible for accepting the Pre-Apprenticeship students and enabling their further training. As mentioned above, there is a lack of skilled tradespersons to meet industry needs. Industry perceptions regarding the Pre-Apprenticeship Program and recognition of it has the opportunity to significantly influence its success.

**Pre-Apprenticeship Students (Parents/ Guardians).** The students are the primary focus of the Pre-Apprenticeship Program. They may, in fact, come to the program for a variety of reasons. However, it is expected that they primarily are looking for an avenue into an apprenticeship and employment. They are seeking personal gain from enrolment in the program. The extent to which they feel they gain from program involvement is important to the program process. Without the students there is no program.

Similarly, as many of the Pre-Apprenticeship students are under the age of 18, their parents or guardians have an interest in the education and training programs in which they are involved. If they do not feel the experience is useful to the teenager for whom they are responsible, or is not beneficial, they may well guide them into another training or educational opportunity.

## **PRE-APPRENTICESHIP PROGRAM THEORY**

### **PROGRAM CAUSATIVE THEORY**

The traditional or intended Pre-Apprenticeship was to provide students with competencies and skills to lead them into further employment or training with advanced standing (NCVER, 2003). This is the basic causative theory underpinning the program as originally intended.

### **PROGRAM IMPLEMENTATION THEORY**

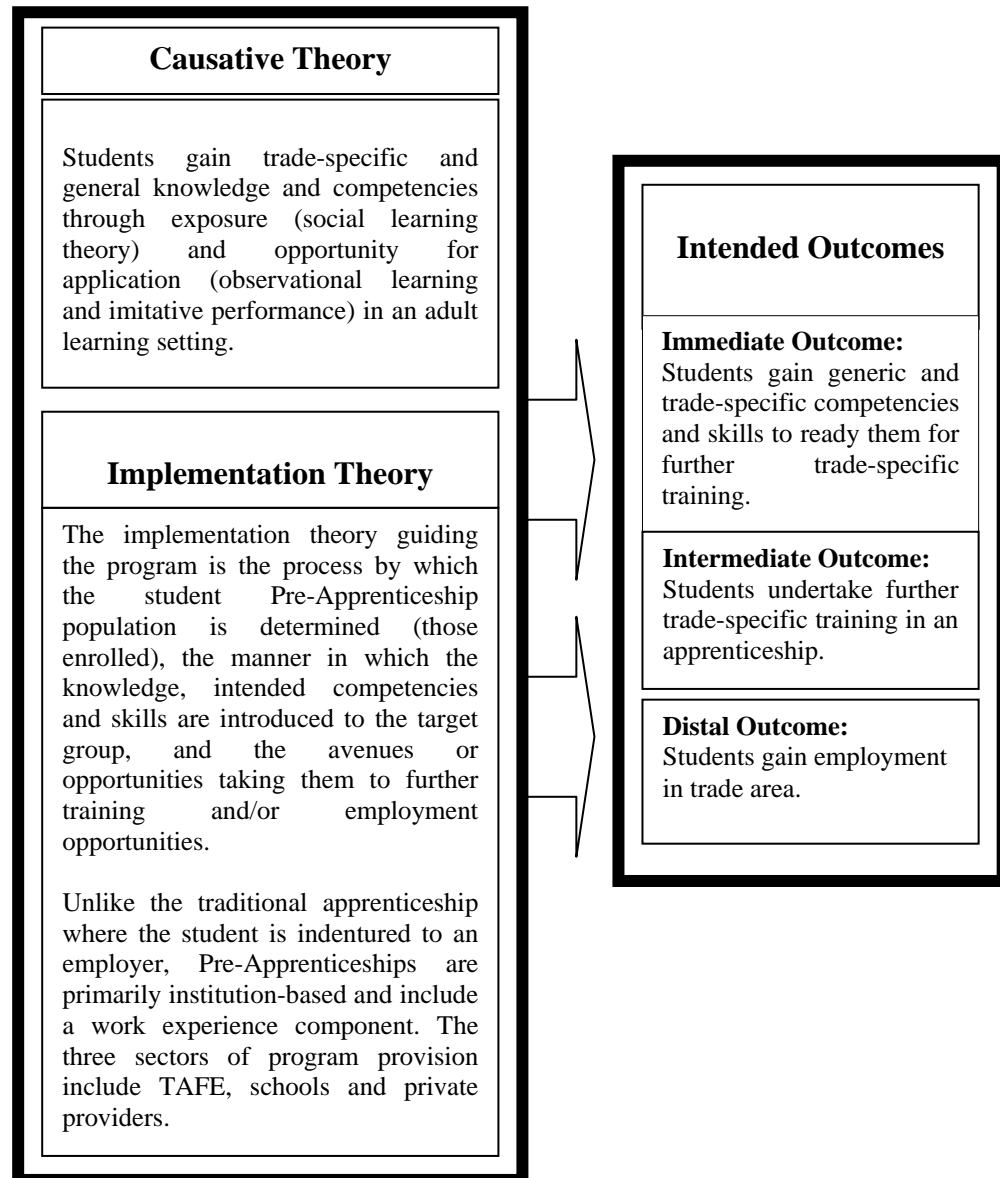
The implementation theory guiding the program is the process of introducing the intended competencies and skills to the target group and taking them into further employment or training.

The 1999 review of Pre-Apprenticeships by the Workplace Skills Management Unit and the August 2003 statistical analysis of the program (Department of Training and Employment, 2003) seemed to indicate that the program had evolved into a form other than that originally intended. The recommendations implemented in 2001 in response to the 1999 review seem to indicate the Department of Training was at that time trying to modify the program implementation to bring it into line with the theory of the program as originally intended. The 2003 investigation of the program again identified gaps in the program implementation (e.g. mismatch between increased number of course enrolments and static apprentice numbers) indicating the intended implementation theory of the program was not what was being actualised.

The available information indicates Pre-Apprenticeship providers have contributed to the shift of the program implementation from the theory of delivery as originally intended. Informal feedback from TAFE staff to the Department of Education and Training indicates the role they see Pre-Apprenticeships playing may also include a vocational introduction and/or a quasi-labour market program. In other words, the causative theory they have adopted for the program is alternative to that originally intended.



The extent to which program implementation was shifted to meet evolving changes in causative theory, or causative theory was adapted to meet changes in implementation theory, is unclear.



**Figure 6.2: Pre-Apprenticeship Program Theory Representation**

## PRE-APPRENTICESHIP EVALUATION

The recommendations of the August 2003 report mentioned above, together with the current political focus on post-compulsory school age education, training or employment related activities, led to the undertaking of the evaluation, with particular consideration of the role of Pre-Apprenticeships in

the system from school to work (Department of Education and Training, 2004). In addition to the traditional intended role of Pre-Apprenticeships outlined above, some argue that Pre-Apprenticeships currently exist to fulfil other roles such as a screening mechanism for employers selecting an apprentice, a quasi-labour market program or vocational introduction providing generic skills making the student ready for work, and a brokerage service linking employers and suitable students as potential employees (Zapelli et al., 2004, September).

The purpose of the Pre-Apprenticeship Program evaluation (Zapelli et al., 2004, September) was to:

- i. Identify and clarify the purpose of Pre-Apprenticeships with particular focus on the roles, benefits and limitations of the three sectors of provision (TAFE, schools and private providers (RTOs)), and the impact on each that would result from changes to the amount of Pre-Apprenticeship delivery;
- ii. Consider the place of Pre-Apprenticeships as part of a system of school to work transition;
- iii. Determine the reasons students undertake Pre-Apprenticeships, the outcomes they achieve with a comparison across the three sectors of provision, and the extent to which client expectations are realised;
- iv. Consider the resources providers have invested in their delivery and assess the cost effectiveness of Pre-Apprenticeships; and,
- v. Determine and investigate the response of Pre-Apprenticeships to the shifts in demand for apprenticeships.

#### PRE-APPRENTICESHIP EVALUATION THEORY-USE CLASSIFICATION

The classification of the Pre-Apprenticeship Program undertaken below was determined after reading the final evaluation report (Zapelli et al., 2004, September), and after interviews with the primary evaluator and many of the key program stakeholders within the Department of Education and Training.

The Pre-Apprenticeship Program evaluation is heavily implementation theory-based, as opposed to causative theory-based. No references have been made in the report detailing the attainment or assessment of particular competencies or

skills, in other words, the underlying causal mechanisms that mediate the relationship between the treatment or course and outcome. No information is contained in the report regarding these mechanisms, nor were they referred to in any of the interviews. Therefore, a rating of *NM* has been given for all three types of theory source under the *Program Causative Theory* section.

Stakeholders were primarily interested what was happening in terms of how the program was being implemented, the expectations, the reality and the outcomes. Implementation theory specifies how the program is put into action. If the program is conducted as planned and the desired outcomes occur, the implementation theory is supported. The Pre-Apprenticeship Program evaluation report focuses on the implementation of the program.

The theory of program implementation outlined at the beginning of the report draws heavily from a review of national and State reports that incorporate relevant social science theory in outlining the role of Pre-Apprenticeships in the vocational training environment. As some of the reports referenced are from the Western Australian Department of Education and Training, some contextualisation is thought to have been considered. Therefore, a rating of *C1* has been given in the *Complex Use* column of *Social Science Theory* under *Implementation Theory*.

Stakeholder Theory of program implementation has been accessed at four levels by the evaluators. The evaluators met with stakeholders from the Department of Education and Training (DET), the body responsible for funding the Pre-Apprenticeship Program. They accessed relevant DET data, records, and documentation in undertaking the evaluation. The evaluators interviewed stakeholders with influence on program delivery from the three sectors of provision (TAFE, schools and private providers). They also interviewed and surveyed those enrolled in the program and the representatives of industries within which the pre-apprentices aim to secure apprenticeships and employment. Therefore, the rating for *Stakeholder Theory* is *Complex* with a contextual rating of *C3*.

Finally, Evaluator Theory is considered to have guided this study. Although Stakeholder theory and Social Science theory have been accessed through a variety of sources, the task of tying it into a coherent statement has been left to the evaluators. In fact, in interviews with DET staff it was said they had endeavoured to not pass along to the evaluators a sense of what they thought the program was, as they were most interested to see what the evaluators found the program to be, and did not want to bias them.

In their consideration of program implementation, the evaluators identified the socio-political environment surrounding the program and considered the setting of the program within the State's vocational and training strategies for young people. The current policy focus on education, training and employment of those of post-compulsory school age and the need to consider the place of Pre-Apprenticeships in the broader system of school to work transition, are mentioned as catalysts for the evaluation. In the case of the Pre-Apprenticeship Program, the Evaluator Theory is considered to be complex and the consideration of context also *Complex* (C3). The classification of the Pre-Apprenticeship Program Evaluation theory use is presented below in Table 6.2.

#### MAKING CONSISTENT JUDGMENTS PROFESSIONAL LEARNING MODULE

#### **PROFESSIONAL LEARNING FOR SCHOOL LEADERS AND TEACHERS**

A key focus of the Western Australian Government's Plan for Government Schools 2004-2007 is the development of teachers and school leaders who are capable and motivated to undertake their work in an environment where they perceive their work to be valued and supported. A primary objective of the plan is the promotion of professional knowledge, learning and expertise with a view to enhancing staff capabilities and improving the status of the teaching profession as a whole (Department of Education and Training, 2004, June). This objective is a key focus of the Curriculum Directorate Plan supporting the implementation of the Professional Learning Program for school leaders and teachers.

**Table 6.2: Theory-Use Classification of the Pre-Apprenticeship Evaluation**

	Program Implementation Theory <i>Program Theory Source</i>									Program Causative Theory <i>Program Theory Source</i>								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial
Program Evaluation Study																		
Pre Apprenticeship			C1			C3			C3	NM			NM			NM		
<p>Note: NM in a box indicates the theory origin was not mentioned in the information sources included in the investigation.  C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based.  The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).</p>																		

The Professional Learning Program encompasses five modules and other elective units for implementation to support Phase Two of the Curriculum Improvement Program (CIP 2). These modules are:

- i. Leading Curriculum Improvement;
- ii. Making Consistent Judgments;
- iii. Planning for Progress;
- iv. Practical Pedagogies; and
- v. Communicating with Parents.

The first two modules of the program were delivered in 2004. It is the evaluation of the second module, Making Consistent Judgments (MCJ) that is the focus of this study.

The model for the CIP 2 Professional Learning Program is underpinned by a Statement of Professional Learning (Department of Education and Training, 2004). The Statement refers to the conceptualisation of effective approaches to professional learning among local, national and international researchers, relevant specialists in the field, and key policy makers. Consensus indicates the most powerful models of professional learning are program-based, systemic in nature, and are underpinned by:

- i. the active involvement of staff in the negotiation and review of their learning;
- ii. the acknowledgement that many staff already have considerable knowledge and experience which can be built upon and prepares them for new learning; and
- iii. the linking and embedding of new learning in daily practice undertaken in a work environment characterised by supportive collegiate teams, and opportunities to apply their knowledge and reflect upon the process and outcomes.

Approaches acknowledged to yield the most effective professional learning include:

- i. connections to both local needs and organisation strategies and initiatives;
- ii. sequenced, spaced learning opportunities to undertake collaborative problem-solving tasks in their daily work setting and given both opportunity and support to reflect on the process;
- iii. support mechanisms to sustain learning;
- iv. flexible access and delivery modes;
- v. deliveries by highly trained and knowledgeable facilitators who value the knowledge and expertise staff already have;
- vi. a range of support materials to cater for different styles of learning;
- vii. evaluation to assess impact and inform further program planning; and
- viii. formal articulation with training and academic accreditation and further career opportunities.

The intention of the CIP 2 is to facilitate coherence between curriculum, pedagogy, assessment, reporting, the learning environment and standards. Key elements of the program include:

- i. a review of reporting to parents and policy simplification;
- ii. the use of the standards and the Outcomes and Standards Framework for monitoring and planning for student improvement;
- iii. the improvement of pedagogy and enhancing the quality of teaching; and
- iv. an increased strategic and systemic focus on professional learning for teachers and school leaders based on the needs of staff, school and the system.

### **MAKING CONSISTENT JUDGMENTS PROFESSIONAL LEARNING MODULE**

The Making Consistent Judgments module consists of three one-day workshops and a spaced learning component. Teachers of Year 3 and Year 9 English and Mathematics are the target audience of the module. Nearly 5000 school staff are expected to be involved in the module from schools across the State. The modules are delivered in the districts by facilitators who are centrally trained.

The aim of the module is to make the Year 3 and Year 9 standards for Mathematics and English explicit and support teachers to implement a moderation process endorsed by the system, to enable the making of consistent judgments against students' achievements against Year 3 and Year 9 English and Mathematics standards.

The ability of teachers to make consistent and comparable judgments about student achievement is necessary to:

- i. improve and refine pedagogy;
- ii. inform and focus school improvement strategies;
- iii. facilitate the collection and reporting of system level information to inform system policy and planning; and
- iv. enable comparative student performance information.

In terms of the Outcomes and Standards Framework, teacher consistency of judgments regarding student achievement rests on the shared interpretation of learning outcomes and standards, and a common understanding of how the achievement of the outcomes and standards are achieved. Moderation of students' work enables a comparison of judgments with colleagues to intricately understand students' achievements. Moderation is an avenue by which teachers' judgments may be validated, or the teacher's understanding of a student's performance is moderated (Department of Education and Training, 2004, June).

### **MAKING CONSISTENT JUDGMENTS PRIMARY STAKEHOLDERS**

The primary stakeholders of the program are considered individually below.

**The Department of Education and Training.** The Department of Education and Training (DET) is responsible for the formal, comprehensive education of students throughout Western Australia through 770 schools. The mission of DET is to promote the development of students who will make a positive and significant contribution to society.

The **Curriculum Directorate** is a division of DET. The Curriculum Directorate is responsible for the improvement of the Western Australian school curriculum for kindergarten through to Year 12 and professional learning. As discussed above, the MCJ module is one module of a professional development package developed and implemented through the Curriculum Directorate with funds and support from the Australian Government Quality Teacher Program (AGQTP). The Directorate is accountable to the Australian Government for effective use of the funds. The Directorate also has the responsibility of professional development and curriculum improvement for the State of Western Australia. They are the primary stakeholders of the program.

**The Australian Government Quality Teacher Program (AGQTP).** The AGQTP is the primary avenue for the implementation of the *Teachers for the*



21<sup>st</sup> Century initiative of the Australian Government. The initiative is directed at:

- i. improving teacher quality;
- ii. increasing the number of effective Australian schools; and
- iii. maximizing student outcomes.

The AGQTP was an election commitment made in 1998 by the Commonwealth Government aimed at updating and improvement of both teachers' skills and the understanding of the professional status of teaching. In the 1999-2000 Commonwealth Budget \$77.7 million was allocated over three years for the program, with \$76.197 million of that amount available for program activities. In the 2002-2003 Commonwealth Budget an additional \$82.4 million was extended to the program and the program provided to June 2005 ([http://www.qualityteaching.dest.gov.au/Content/Item\\_1115.htm](http://www.qualityteaching.dest.gov.au/Content/Item_1115.htm); 17/06/05). The AGQTP has three components. They are:

- i. the provision of State/Territory projects aimed at professional learning in priority areas;
- ii. national initiatives involving a range of projects including a focus on teachers, teaching and school leadership related issues; and
- iii. the National Institute for Quality Teaching and School Leadership and responses of the Australian Government to the Lee Dow Review of Teaching and Teacher Education.

The MCJ module implementation is funded by the AGQTP. A requirement of that funding is that the program developed and delivered is evaluated and the use of the funds is substantiated. Therefore, the Australian Government, the stakeholders of the AGQTP, is a key stakeholder of the MCJ module and its evaluation.

**Curriculum Council.** The mission of the Curriculum Council is to determine the direction of Western Australian curriculum policy for kindergarten through Year 12 schooling. In doing so, the Curriculum Council supports the development and implementation of a Curriculum Framework to guide schooling that incorporates: students' needs; the determination of the knowledge, comprehension, skills, values and attitudes students are expected

to acquire; and the assessment and certification of student achievement. The Curriculum Council has involvement in the development and accreditation of post-compulsory schooling courses. As the Curriculum Council has stated on their web page, “Student learning is at the heart of everything we do” (<http://www.curriculum.wa.edu.au/pages/council/council00.htm>; 16/06/05).

The Curriculum Council has an interest in the MCJ module as it potentially impacts upon the learning of Western Australian students. However, more specifically, the MCJ module focus is on improving the consistency of teacher judgment of student achievement in terms of the Outcomes and Standards Framework for monitoring and planning for student improvement. Although the Curriculum Council have not been directly involved in the MCJ module implementation or its evaluation, it has a keen interest and clear stake in both.

**Western Australian Primary Principals’ Association (WAPPA).** WAPPA was set up to promote the professional expertise and status of primary school principals and enhance Western Australian primary school education. The primary objective of WAPPA is to foster both educational thought and practice of educators and establish a high professional standard (<http://www.wappa.asn.au>; 16/06/05). It clearly has an interest in any professional development undertaken by primary school staff. Although a WAPPA representative did not sit in on the meetings to initially conceptualise the evaluation, an interview with the WAPPA President found that the organisation keenly monitors the progress of the MCJ module and is very interested in the information that results from its evaluation.

**Western Australian Secondary Schools Association (WASSA).** The primary objective of WASSA is to promote the professional development and ability of secondary school executives and to enhance the quality of secondary school education in Western Australia (<http://members.iinet.net.au/~wassea/about.html>; 16/06/05). In their work towards this objective, WASSA promotes the professional interests of its members by:

- i. organising and conducting regular meetings, forums, conferences, etc, to further the professional education of members;
- ii. disseminating professional information to members;
- iii. instigating and advocating education research and study;
- iv. acting as a recognised consultative group to the Minister for Education and other key agencies; and
- v. representing members in negotiations.

WASSA is responsible for the professional development of its members and to carefully monitor any undertaking which might impact upon the quality of secondary school education. Therefore, it has closely monitored the MCJ module implementation and is very interested in any information that results from the module's evaluation.

**Teachers of Western Australia.** The teachers of Western Australia have a stake in the MCJ module. The extent to which they take the professional learning on board and apply it to their work setting are clear intended outcomes of the program. Teachers invest time and energy with the expectation they will gain knowledge and skills which can be applied to their work and enable them to be more informed and more effective educators.

**Students and Parents (Guardians).** Students are the distal focus of the program. The professional learning modules have been developed and implemented with a view to enabling teachers to better educate the students of Western Australia. Students, and the guardians responsible for their care, have a stake in the MCJ module and its evaluation, as the extent to which the program is successful ultimately impacts upon them.

## **MAKING CONSISTENT JUDGMENTS PROGRAM THEORY**

### **CAUSATIVE PROGRAM THEORY**

The broad causative theory underpinning the program states that if a professional learning program is implemented in accordance with the principles outlined above, professional learning will be enhanced and more

likely to be implemented on a long-term basis in the participants' place of work. The causative theory underpinning the program focuses not only on the acquisition of knowledge and skills by staff, but also on the practical use of the knowledge in a work setting. The full impact of the causative theory, as intended, would indicate distal as opposed to immediate effect. Assessment of the impact of the program on professional learning, through feedback surveys at the end of the workshop, might not capture the true picture regarding the extent to which the program has influenced professional learning.

#### **IMPLEMENTATION PROGRAM THEORY**

The key principles guiding the professional learning program are primarily regarding implementation. These include conditions such as: flexible module delivery; highly trained presenters; the availability of resource materials catering for a variety of learning styles; the acknowledgment of the knowledge and skills already held by staff undertaking the course; the inclusion of collaborative learning opportunities to be taken back to their work setting; support in taking the knowledge gained in undertaking the course back to the work setting; and opportunities to reflect on workplace trials within the course and gain constructive feedback. The model adopted for the Professional Learning Program is very different to the way professional learning has been delivered by the Department of Education and Training in the past. The logic guiding the adoption of the alternative mode of delivery rests upon the theory that professional learning delivered in such a way will have a greater impact, in terms of increasing the knowledge and skills of staff, as well as facilitate the transfer of the knowledge and skills gained to the work environment.

#### **MAKING CONSISTENT JUDGMENTS PROFESSIONAL LEARNING MODULE EVALUATION**

The purpose of the evaluation is to assess the extent to which the module provides professional learning and support to school leaders and teachers in accordance with key performance indicators (Cummings et al., 2005, June). The evaluation team held a number of collaborative meetings with key Department of Education and Training (DET) staff to clarify the key

performance indicators, develop evaluation questions to be addressed and discuss information collection strategies. Questionnaires prepared by DET to collect qualitative and quantitative feedback information are the primary source of data collection. Key performance indicators are the extent to which:

- i. the module was relevant to the learning needs of teachers and school leaders;
- ii. the skills and knowledge to teachers and school leaders were enhanced;
- iii. knowledge and expertise was integrated into workplace practice; and
- iv. facilitator training was effective.

As a result of being involved in the module program participants should be able to:

- i. identify where they and their team are regarding the implementation of the CIP 2 and develop an implementation plan for their school;
- ii. determine their needs in terms of skill and understanding to support the implementation process and develop an individual professional learning plan; and
- iii. commit to collegiate support opportunities to support further professional learning.

#### **MAKING CONSISTENT JUDGMENT PROGRAM THEORY USE CLASSIFICATION**

The classification of the Making Consistent Judgments evaluation theory use is based upon a review of the study data collection instruments and consideration of information gained in interviews with key program and evaluation stakeholders. The primary evaluator led the evaluation reference group through a process considering program outcomes, performance indicators and instrument development early in the evaluation conceptualisation. The classification below rests primarily on a review of the evaluation instruments developed in that process.

Considerable social science theory and stakeholder theory underpin the Professional Learning Program and, in turn, the CIP 2 implementation and Making Consistent Judgments Module. Program implementation information has been gathered mainly from surveys of the module presenters. However, some items regarding module implementation have also been included in the

staff survey. The items focus on issues pertinent to the program implementation theory outlined above. As the items have been developed in negotiations between the evaluators and the program stakeholders; program stakeholders well informed regarding the social science theory underpinning the program, all three theory sources (Social Science Theory, Stakeholder Theory and Evaluator Theory) have been accessed in the evaluation's consideration of program implementation theory. However, the items included in the surveys are not particularly complex. Social Science Theory, Stakeholder Theory and Evaluator Theory have all three been accessed in a *Simple* manner in the evaluation process.

Similarly, the consideration of context in the development of the survey items has not been particularly detailed, and is most evident in the initial items of the program registration form which focuses on the details of the school from which the staff member has come. A rating of (C2) has been given to rate the contextual consideration with regards to the program implementation theory of the stakeholders and the evaluators upon which the evaluation is based. Contextualisation of Social Science Theory is a product of the stakeholders and evaluators conceptualisation of program theory. A rating of NM has been given for Social Science Theory consideration of program context.

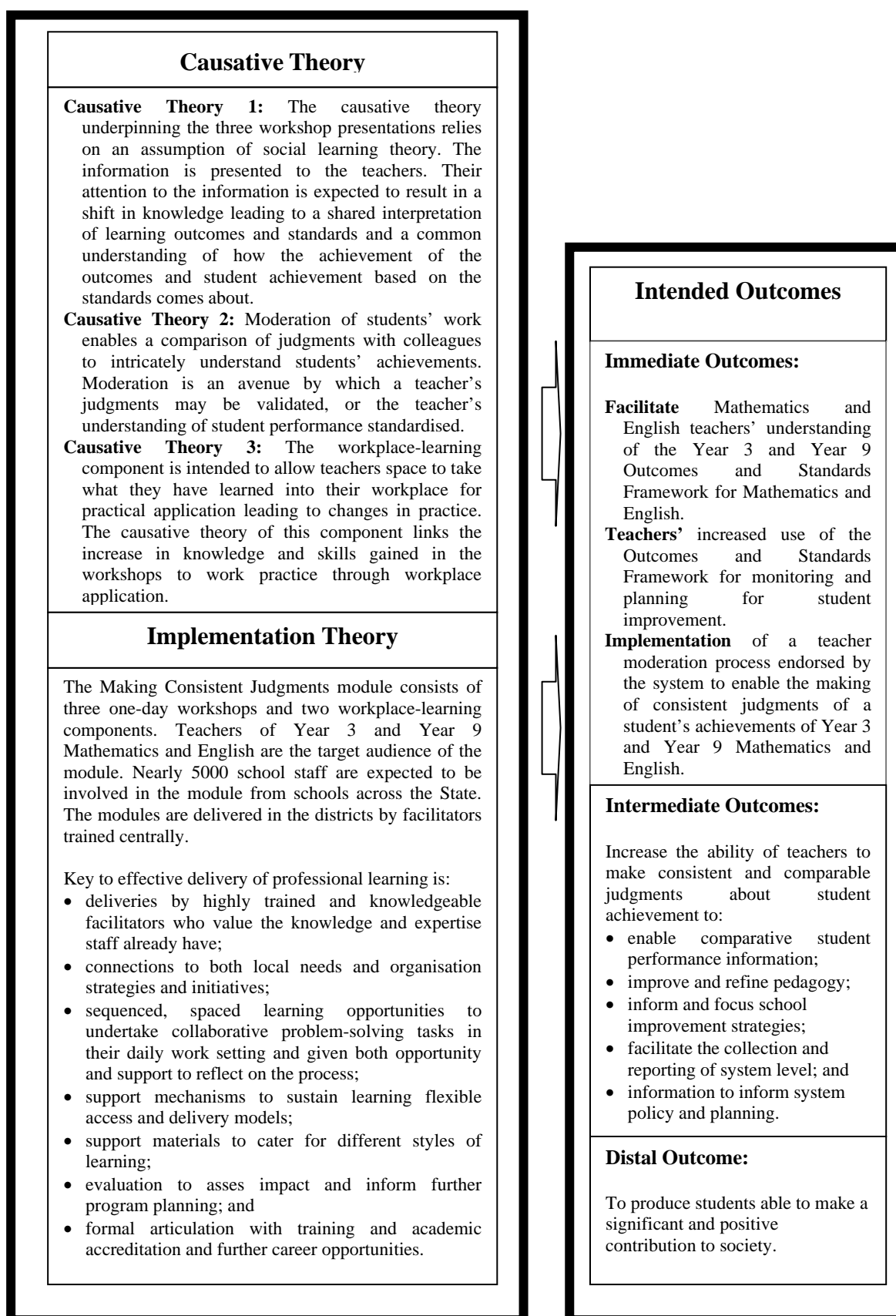
The evaluation's consideration of program causative theory is a little more complex. Survey items focus on the knowledge and skills gained by staff as a result of attending the MCJ module, and on changes in behaviour intended to result from module participation. As the stakeholders and the evaluators together negotiated the items, the influence of both stakeholder and evaluator theory is considered to be complex. As social science theory underpins the program theory conceptualisation of stakeholders, it has also been used in a complex manner in the evaluation process.

Contextualisation of the Causative Program Theory used in the evaluation is Simple (C2) with regards to Stakeholder Theory and Evaluator Theory. A rating of NM has been given for Social Science Theory consideration of

program context. The theory-use classification of the Making Consistent Judgments evaluation is summarised in Table 6.3.

**Table 6.3: Theory-Use Classification of the Making Consistent Judgments Evaluation**

	Program Implementation Theory <i>Program Theory Source</i>									Program Causative Theory <i>Program Theory Source</i>								
	Social Science Theory			Stakeholder Theory			Evaluator Theory			Social Science Theory			Stakeholder Theory			Evaluator Theory		
Program Evaluation Study	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial	No or Negligible	Minimal	Substantial
<i>Making Consistent Judgments Evaluation</i>		NM			C2			C2				NM			C2			C2
<p>Note: NM in a box indicates the theory origin was not mentioned in the information sources included in the investigation.</p> <p>C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based.</p> <p>The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).</p>																		



**Figure 6.3: MCJ Program Theory Representation**



## THE CLASSIFICATION MATRIX

Table 6.4 below represents the theory classification of all three evaluations included in this study to enable the comparison and contrast of theory-use classifications.

**Table 6.4: Theory-Use Classification of the Pre-Apprenticeship, YOHFest and MCJ Program Evaluations**

Evaluation		Implementation Theory			Program Causative Theory		
		Program Theory Source			Program Theory Source		
		<i>Social Science Theory</i>	<i>Stakeholder Theory</i>	<i>Evaluator Theory</i>	<i>Social Science Theory</i>	<i>Stakeholder Theory</i>	<i>Evaluator Theory</i>
Pre-Apprenticeship Program Evaluation	No / Negligible				NM	NM	NM
	Minimal						
	Substantial	C1	C3	C3			
YOHFest Program Evaluation	No / Negligible	NM	NM	NM			
	Minimal						
	Substantial				C1	C1	C1
MCJ Program Evaluation	No / Negligible						
	Minimal	NM	C2	C2			
	Substantial				NM	C2	C2
	<p><b>Note:</b> NM in a box indicates the theory origin was not mentioned in the information sources included in the investigation.  C denotes the consideration of program context in the elaboration of Program Theory upon which the evaluation is based.  The number accompanying the C denotes the level of contextual consideration; No or Negligible (1), Minimal (2), or Substantial (3).</p>						

## MEASURE OF PROGRAM THEORY USE CLASSIFICATIONS

The final step in the classification of type, degree and contextualisation of program theory use adopted in each case, is the determination of numerical measures based on the Classification Matrix data for each evaluation. The first portion of this section summarises the logic underpinning the values assigned to designate program theory use by type, degree and level of contextualisation (see Chapter II for more detail). The remainder of this section deals with the values assigned to each evaluation of this study regarding the use of program theory.

### PROGRAM THEORY USE CLASSIFICATION SCALING

In the selection of the evaluation studies an attempt was made to include studies which varied in the degree of use of program theory with regards to program implementation theory and program causative theory. In keeping with the original design proposed for this study it was decided that two scores would be given for program theory use for each case, one for program implementation theory use and one for causative theory use.

A score of '0' to '6' is possible for each source of program theory (Social Science Theory, Evaluator Theory, and Stakeholder Theory) within each type of theory use (Implementation & Causative Program Theory). Simply, a score of '0' has been assigned for 'No or Negligible' theory use. A score of '1', '2' or '3' is possible for 'Minimal' program theory use, depending on the level of contextual consideration included in the theory (C1, C2, or C3), while a score of '4', '5' or '6' is possible for 'Substantial' program theory use, again depending on the level of contextual consideration. The scoring system is represented in Table 6.5 below. The theory use scores for the three cases of this study are represented in Table 6.6.

**Table 6.5      Program Theory Use Score Pattern**

Program Theory Use Score		
Level of Program Theory Use	Level of Contextual Consideration	Score
<i>No/Minimal</i>	N/A	0
<i>Simple</i>	C1	1
	C2	2
	C3	3
<i>Complex</i>	C1	4
	C2	5
	C3	6

**Table 6.6      Program Theory Use Scores for Case Study Evaluations**

	Implementation Theory <i>Theory Source</i>				Causative Theory <i>Theory Source</i>			
	Social Science Theory	Evaluator's Theory	Stakeholder's Theory	Implementation Theory Use Score Total	Social Science Theory	Evaluator's Theory	Stakeholder's Theory	Causative Theory Use Score Total
<b>EVALUATION</b>								
<b>YohFest Program</b>	0	0	0	<b>0</b>	4	4	4	<b>12</b>
<b>Pre-Apprenticeship Program</b>	4	6	6	<b>16</b>	0	0	0	<b>0</b>
<b>Making Consistent Judgments Program</b>	0	2	2	<b>4</b>	0	5	5	<b>10</b>

The resulting scores represented in Table 6.6 indicate the YOHFest Program evaluation was underpinned by heavy use of the program's causative theory (score '16') and no program implementation theory (score '0'), whereas the Pre-Apprenticeship Program evaluation was driven by implementation theory (score '16') and no program causative theory (score '0'). Finally, the Making Consistent Judgments Program evaluation was underpinned by some use of the program's implementation theory (score '4') and heavier use of the program's causative theory (score '10'). The scores outlined above indicate a range of scores has been achieved.

### SUMMARY

The sections of this chapter have included detailed descriptions of the programs evaluated in this study, including the programs' aims, primary stakeholders of the programs, and programs' theory focusing on both the implementation and causative portions. At the end of each section dealing with a program, a classification of the program's theory is represented. The final sections of this chapter focus on the classification of the use of program theory in the evaluations of the programs. A method for scoring program theory use by type and degree is presented and scores assigned to each evaluation regarding program theory use.

## CHAPTER VII

### FINDINGS OF THE STUDY

One of the primary purposes of this study was to develop and analyse a model of factors influencing the use of a program evaluation and the information it yields. The model focused on a set of factors identified in previous research as potentially influencing the use of information produced by evaluation studies, including the characteristics of the organisation responsible for the delivery of the program, the stakeholders of the evaluation, the evaluator, and characteristics of the evaluation study. Of particular interest was the influence on utilisation of the use of program theory in the planning and conduct of an evaluation study, a factor not previously investigated. The model was tested using the data from a longitudinal study of three program evaluation studies. The analysis focused on descriptive information about each program, the evaluation study conducted for each program and a series of interviews with key stakeholders and the evaluators.

The focus of this chapter is on reporting the findings of the study. The first section concentrates on the structural equation modelling (SEM) process, including the influence of the parameters, estimation and sample size. The second section focuses on the development of the measurement model for the study, including details of the model fit assessments for each of the four measurement sub-models of the study. The structural model of the study is the focus of the third section of this chapter and the results of the structural analysis are provided. Analysis of **Moment Structures**, or AMOS version 7.0 (Arbuckle, 1995-2005), is the structural equation modelling computer package used to conduct the analysis of the model for this study.

### STRUCTURAL EQUATION MODELING

The initial focus of this section is on the logic underpinning SEM, including the merits of using SEM to analyse data, the pertinent terminology, and

consideration of the two components of SEM, the measurement model and the structural model. Structural equation modelling was chosen due to the increased capacity of the modelling process to accommodate the interplay between theory and data. SEM allows the construction of unobserved (latent) variables from observed variables, the modelling of relationships among multiple exogenous (independent) and endogenous (dependent) variables, the calculation of errors of measurement for the observed variables, and statistical testing of both theoretical and measurement assumptions using empirical data. SEM is based on an extension of first-generation statistical procedures such as multiple regression and factor analysis, yet enables the researcher to account for measurement error, often sizeable in social research methods (Byrne, 1989; Cummings, 1997; Holmes-Smith, 2001).

Although this data was hierarchical in nature (e.g. each evaluation had a stakeholder group, an evaluation team and organisational representatives) a hierarchical level model structure was considered. However, a model-based approach was chosen to analyse the data. The SEM approach was adopted by Cummings (1997) in his longitudinal case study design adopted for this investigation. The data was structured in such a way that each individual had a string of data. Structuring of the data in this way enabled good insight into perceptions of use of the evaluations and their information.

SEM has merit when studies of experimental design are not appropriate due to ethical considerations, or methods for testing particular theories are not well developed. It is a statistical methodology which applies a confirmatory approach to the examination of a theory structure modelling some phenomenon, usually representing causal processes. SEM has many qualities not found in other multivariate procedures that were useful for the analysis of the data of the present study (Byrne, 1989; Cummings, 1997; Holmes-Smith, 2001). Firstly, it takes a confirmatory rather than an exploratory approach to data analysis. Second, as SEM requires the specification of a model of variable inter-relationships prior to statistical analysis, the analysis of data for inferential purposes is possible, unlike many other multivariate procedures which are predominantly descriptive, such as exploratory factor analysis.

Third, through the use of SEM, researchers have the ability to examine more complex relationships and models, including testing a set of regression equations simultaneously. Fourth, SEM explicitly estimates error variance parameters, a function more traditional multivariate procedures are unable to do. For example, methods based on regression analysis or the general linear square model, assume the error in the independent variable disappears. Finally, other multivariate methods are strictly based on observed measures. However, SEM can include both unobserved latent variables constructed from observed variables, and observed variables (Bryne, 2001; Goldberger, 1973; Hoyle, 1995; Information Technology Services at the University of Texas at Austin, 2002).

Knowledge of the terminology used in SEM is necessary to understanding it. Manifest or observed variables are directly measured by researchers. Latent, unobserved variables are not directly measured, but are inferred by the relationships or correlations among observed variables in the analysis. The statistical estimation undertaken in SEM is accomplished in much the same way that an exploratory factor analysis infers the presence of latent factors from shared variance among observed variables (Bryne, 2001).

In general, the SEM process is undertaken in two stages. First, the measurement model is developed which specifies the relationships between the observed variables and latent variables, linking the scores of the measurement instruments employed for data collection and the latent or underlying constructs they have been developed to measure. The measurement model identifies the reliabilities and validities of these relationships and is independent of the structural relationship between the latent variables.

The second stage, analysis of the Structural Model, involves the analysis of the relationships among the latent variables. It defines the manner by which these variables influence each other in the model either directly or indirectly. Influence is determined by the related in the values of other latent variables in the model (Bryne, 2001; Cummings, 1997).

## MODEL FIT

Parameters, estimated from the data, are essential to the interpretation of both the measurement model and the structural model. Parameters are the regression coefficients and the variances and co-variances of the variables. For the purpose of testing and comparing parameter estimates, unstandardised estimates are used. Unstandardised estimates retain the scaling information of the variables used and are interpreted with reference to the variable scales. Alternatively, standardised parameter estimates are based on mean comparison procedures (e.g. t-test and ANOVA), and based on transformations of the data to standardize these data to a common scale.

A number of conditions regarding the parameters of the model must be met before it can be identified. A necessary condition is that the number of free parameters ( $q$ ) is equal to, or smaller than, the number of non-redundant parameters, generally known as  $p$ , which is equal to  $p(p+1)/2$  (Chou and Bentler, 1995). Once it has been determined the model is *identified*, the parameters of the model are estimated from the data and the adequacy of the model determined.

### METHOD OF ESTIMATION OF MODEL PARAMETERS

Regression coefficients and the variances and co-variances of independent variables are estimated by means of the goodness-of-fit indices. Goodness-of-fit statistics indicate the similarity between the co-variance matrix of the estimated model and the population co-variance matrix. Interpretation of the model is based on these statistics. However, they must be estimated from the data. Therefore, estimation, after model specification, is the first step in the SEM modelling process. Goodness-of-fit test statistics are calculated at the same time as the estimation and are dependent on the method of estimation selected.

Maximum Likelihood (ML) is the estimation method most commonly used in SEM. Four primary assumptions underpinning the statistical properties of the



ML estimation procedure are that: the sample is very large (asymptotic); the observed variables are continuous; the observed variables have a multivariate normal distribution; and the model estimated is valid.

Hoyle and Panther (1995) argue that there is a significant body of research which indicates ML statistics derived when the data is not optimal, such as small sample size or excessive kurtosis, are nevertheless reasonably good. Research indicates ML tends to be robust to violations of normality, in that the estimates tend to be good even in cases where the data are not normally distributed. Furthermore, when both common and error variances are distributed independently of each other, all fit indices were less consistent with ML, GLS (Generalised Least Squares) and ADF (Asymptotically distribution-free) when samples sizes were small. However, West, Finch and Curran (1995) recommend the ML or GLS estimates for small sample sizes when the distributions are substantially non-normal.

The assumption that estimates and tests are based on large samples underpins all estimation procedures, but this is often not the case in practice. However, the accuracy of the estimation is open to the influence of sample size, with small samples likely to bring about the least accurate results (Holmes-Smith, 2001; Hu and Bentler, 1995; MacCallum, 1995). Hoyle and Panther (1995) recommend ML estimation be used as it is the most widely available estimator, unless characteristics of the data indicate it not the most appropriate choice. For the purposes of this investigation, ML is the estimation method employed.

The question of how large a sample should be so as to enable estimation and testing of structural equation models with latent variables, is a focus of many authors (Bentler and Chou, 1987; Holmes-Smith, 2001; Stevens, 1996; Marsh, Balla and McDonald, 1988). The potential effects of small samples include increased frequency of convergence failures (i.e. AMOS cannot reach a satisfactory solution), improper solutions such as negative error variance estimates for measured variables, and reduced accuracy of parameter estimates. The accuracy in determination of standard errors is reduced in

particular, as it is computed under the assumption of large sample sizes (Information Technology Services at the University of Texas at Austin, 2002).

Thresholds for sample size have been suggested by a number of authors. For instance, Bentler and Chou (1987) recommend five cases per parameter estimate. Alternatively, Stevens (1996) recommends 15 cases per predictor in a standard ordinary least squares multiple regression analysis. It has been argued five cases per parameter and 15 cases per predictor actually imply similar sample size requirements (Holmes-Smith, 2001).

The small sample size (36 subjects) of the present study has presented some difficulties in the structural equation modelling process and has been taken into account in the assessment of model fit. A larger sample size would have encouraged more confidence in the evaluation of the model fit and enabled an increased degree of flexibility in testing the model (e.g. split half analysis). However, it may be argued with regards to the current study that as each respondent is observed at three data points, the sample size might be considered less deficient than the actual number of subjects indicates. Furthermore, as with the work of Cummings (1997) upon which a substantial portion of this study is based, the internal consistency of the scales (inter-item reliability ( $r_{ii'}$ ) and item total reliability ( $r_{it}$ )) contributing to the latent variables of the model is high (refer to Chapter V of this dissertation for analysis of the scales). Finally, previous research and relevant evaluation literature strongly supports the overall hypothesised model (refer to the literature review in Chapter II and Model Development in Chapter III) of this dissertation, encouraging confidence in the models being investigated. It should also be noted that this investigation represents an initial exploration into a number of areas which have not previously been investigated with a view to further investigation with larger sample sizes.

## **ASSESSMENT OF MODEL FIT**

The assessment of model fit follows a pattern in terms of the order in which the output information is reviewed (Bryne, 2001; Holmes-Smith, 2001). The

purpose of determining model fit is to assess the degree to which a model fits with the data from which it has been estimated. The next step is then to determine any areas of model misfit.

The first step in model evaluation is to check the parameters to ensure they are within accepted limits. The statistical significance of parameter estimates is considered next. The critical ratio representing the parameter estimate divided by its standard error, is the test statistic indicating the significance of the parameter estimate. The critical ratio based on a 0.05 level should be  $> \pm 1.96$ , indicating the null hypothesis that the estimate is equal to zero, can be rejected. In the interest of parsimony, non-significant estimates might best be deleted from the model, with consideration given to the adequacy of sample size in the calculation of the estimate. Furthermore, no variance or co-variance should be a positive definite, meaning one is a perfect predictor of the other.

The second step is to assess the appropriateness of standard error. If the standard error measures are too large or too small (approaching zero), the test statistic for the relevant parameter cannot be determined. However, no specific thresholds of small or large values have been determined (Bentler, 1995; Jöreskog and Sorbom, 1988).

Once the soundness of the parameters has been assessed it is reasonable to consider the fit of the model as a whole. A review of the Chi Square ( $\chi^2$ ), degrees of freedom, and probability ( $p$ ) allow a quick insight into model fit. Other elements to consider in relation to model fit include the logic that underpins the model and goodness-of-fit indices.

Goodness-of-fit indices represent the similarity between the co-variance matrix of the estimated model and the co-variance matrix of the population from which the sample was drawn. If the model is not representative of the sample population co-variance, there is no point in interpreting the model parameters (Chou and Bentler, 1995). A variety of fit statistics have been developed over the years (Hoyle, 1995). For the purposes of this study, four fit

statistics have been selected for determination of the measurement model fit. They are; Chi Square ( $\chi^2$ ), probability ( $p$ ), Minimum Discrepancy (CMIN/df), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). The Adjusted Goodness-of-Fit Index (AGFI) and Parsimony Goodness-of-Fit Index (PGFI) have been included as goodness-of-fit indices for the evaluation of the structural models. The fit indices are further described below in the sections dealing with the measurement model and the structural models.

In AMOS, the output provides two types of information useful for model re-specification, should it be a poorly fitting model. Firstly, information regarding the residual values representing estimates of the number of standard deviations between the observed residuals and the zero residuals that would result if the model were a perfect fit. Residual values  $> 2.58$  are too large (Jöreskog and Sorbom, 1988).

Secondly, AMOS provides a modification index (MI) for each fixed parameter. Modification indices also provide information regarding model misfit and are considered in the judgment of individual aspects of a particular model. This index indicates the expected drop in overall  $\chi^2$  value if the parameter were left free in another analysis of the model. Next to the MI is a value expressing the predicted estimated change for each fixed parameter in the model, in either a positive or negative direction, and is particularly important with regards to any changes to the parameters of the model. A MI value greater than 3.84 is significant at the 0.05 level. In this study only parameters of 4.00 were considered for re-specification. However, many parameters with values of 4.00 or above were not re-specified as there was no substantive reason to do so (e.g. parameters were unrelated). If no parameter estimates are displayed, it means that none exceed the specified threshold.

*Modification indexes and residual co-variances* are referred to in consideration of modifications to the model to increase its fit with the data from which it has been estimated (Bryne, 2001; Information Technology

Services at the University of Texas at Austin, 2002; Holmes-Smith, 2001). The implications of these measures are reported in the assessment of each model as appropriate. If modification indices and residual co-variance values are within normal limits, no modification of the model is likely to be undertaken.

The application of the AMOS output and the interpretation of the statistics are considered further in the assessment of the measurement and structural models below. It is useful, however, to mention Bentler and Chou's (1987) caution that even a model which appears to fit well according to the statistical measures, may be poorly specified. The statistical measures give insight into the fit of the model, not into whether or not it is believable. Therefore, it is clear the evaluation of a model includes theoretical, practical and statistical considerations (Bryne, 2001).

#### TESTING OF THE MEASUREMENT MODEL

The fit indices selected for evaluation of the measurement model include Chi Square ( $\chi^2$ ), minimum discrepancy (CMIN/df), root mean square error of approximation (RMSEA) statistic, and Tucker-Lewis Index (TLI). In assessment of model fit we first consider the degrees of freedom,  $\chi^2$  and the probability of attaining the  $\chi^2$ . If these statistics do not indicate model fit, no further investigation of fit indices is undertaken. Each of the fit indices is discussed below and literature relevant to each summarised.

##### **CHI SQUARE ( $\chi^2$ )**

The  $\chi^2$  statistic is an absolute goodness-of-fit statistic, in that it is concerned with measuring the absolute discrepancy between the matrix of implied variances and co-variances and the matrix of empirical sample variances and co-variances. The  $\chi^2$  is a statistic calculated in relation to the number of degrees of freedom available. In application, the probability ( $p$ ) value of deriving a particular  $\chi^2$  statistic from a set of data determines the significance

of the fit. If the probability value is below 0.05, the departure of the model, from the data is significant at the 0.05 level and the model is rejected. Alternatively, if the  $p$  value is greater than 0.05, the departure of the model from the data is not significant at the 0.05 level and the model is accepted (Bryne, 2001; Hoyle, 1995; Hu and Bentler, 1995).

The sensitivity of the  $\chi^2$  statistic to sample size is important to note as it has been a concern of many authors (e.g. Bentler, 1990; Hu and Bentler, 1995; Jöreskog and Sörbom, 1988), and should be kept in mind when interpreting the statistic. The statistic is derived from a calculation based on a distribution of the sample. If the sample is small, the  $\chi^2$  value is more likely to show the model does not differ significantly from the data. However, if the sample is large the  $\chi^2$  value is more likely to show the model differs significantly from the data, prompting the model to be rejected. The sample sizes of all interview groups included in this study are small, ranging from 10 to 36. Therefore, the tendency of the  $\chi^2$  to more frequently indicate the model is consistent with the data from which it was estimated, is a concern. In order to add confidence, this tendency is avoided by including in this study three other fit statistics to assess the fit of the measurement models. Had the fit of the measurement model been determined by only the  $\chi^2$  value, it is possible models of a poor fit with the data may not have been rejected.

#### **MINIMUM DISCREPANCY (CMIN/df)**

A number of authors suggest that the ratio of minimum discrepancy to degrees of freedom in general, be used as a measure of fit (Wheaton et al., 1977; Carmines and McIver, 1981; Marsh and Hocevar, 1985; Byrne, 1989). The ratio value should be close to one for acceptable models. However, there is some disagreement regarding the thresholds for values above 1, which determine the model as being unsatisfactory. Byrne (1989) concludes a CMIN/df ratio value greater than 2 is indicative of an unsatisfactory model fit. However, Marsh and Hocevar (1985) suggest a ratio value between 2 and 5 to be an acceptable indicator of model fit. Similarly, Wheaton et al. (1977)

recommend a ratio value of 5 or less as acceptable, while Carmines and McIver (1981), suggest acceptable C/MIN ratio values should range between 1 and 3. Finally, Holmes-Smith (2001) suggests values should be greater than 1, but less than 2. However, he considers values between 2 and 3 to be indicative of reasonable fit, while values less than 1 indicate over fit. With regards to the present study, values between 1 and 3 have been accepted to indicate good model fit.

#### **ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA OR RMS)**

The RMSEA is less restrictive than  $\chi^2$  in that it incorporates the error of approximation in the population and does not require the model to hold exactly in the population; an unreasonable assumption given that most empirical research is based on samples of the population (Arbuckle, 1995-2005). Alternatively, the RMSEA informs us as to how well the model would fit the population co-variance matrix if available, given unknown but selected optimal parameter values. RMSEA incorporates no penalty for model complexity. In fact, it will tend to favour a model with many parameters. More recently, the RMSEA has been recognised as one of the most informative statistics in co-variance structure modelling (Byrne 2001). However, Hu and Bentler (1999) warn that the RMSEA over rejects true population models when the sample size is small.

There is some disagreement in the literature regarding threshold levels of RMSEA. Many authors suggest a RMSEA value of 0.05 or less to be indicative of good model fit in relation to the degrees of freedom, although values between 0.05 and 0.08 suggest a reasonable fit (Browne and Cudeck, 1989; Byrne, 1989; Holmes-Smith, 2001), but values over 0.10 are considered to be high by others (e.g. Browne and Cudeck, 1989). Due to the small sample size of this study, and the tendency of RMSEA to over reject models due to small sample size, the upper threshold of 0.10 advocated by Browne and Cudeck (1989) has been used in the evaluation of model fit.

## **TUCKER-LEWIS INDEX**

The Tucker-Lewis Index (TLI), also known as the Bentler-Bonett non-normed fit index (NNFI), is one of a number of adjunct fit indexes developed in reaction to a growing dissatisfaction with the  $\chi^2$  goodness-of-fit adjunct test. Fit indexes were designed to avoid some the problems of sample size and distributional mis-specification in the evaluation of a model (Bentler and Bonett, 1980). The TLI, a ‘non-normed’ fit index pioneered by Bentler and Bonett, (1980), is an extension of the fit index developed by Tucker and Lewis (1973). As with other adjunct fitness indexes, it is a measure of model fit derived from comparison of the fit of a specified model and the fit of an independence model, a model in which no relationships among variables are specified (null model). In the case of the independence model, only the variances are estimated, as all relationship pathways are set to zero (Hu and Bentler, 1995).

Bollen (1986) suggested the TLI to be dependent on sample size. However, as Marsh et al. (1988) point out, Bollen included no rigorous tests or mathematical proofs substantiating his claim. Alternatively, Marsh et al. (1988) examined the effect of sample size on 30 fit indices for both real and simulated data. The TLI was the only widely used index they found to be relatively independent of sample size. In consideration of the small sample size of the present study, the selection of the TLI as an index for model evaluation seemed most appropriate.

Generally, adjunct fit indexes range between 0.00 and 1.00. They are not statistical tests of model fit, but rather indices of model adequacy. TLI values can extend beyond the 0-1 range and is a descriptive fit index interpreted intuitively. A value of 0.90 is widely considered to be the threshold an index must exceed before a model is considered to be consistent with the observed data from which it was estimated (Hoyle, 1995; Byrne, 1995). With regards to the present study, measurement models with TLI values between 0.90 and 1.00 are considered to be a good fit with the data.



## PRIORITY OF FIT STATISTIC CONSIDERATION IN MODEL ASSESSMENT

In the consideration of model fit Bryne (2001) advocates considering the  $\chi^2$  fit statistic as the first step in assessing model fit. If an acceptable  $\chi^2$  value is found, attention is turned to the other three fit statistics, generally in the order CMIN/df, TLI and RMSEA. Once the general fit of the model has been assessed, the implications of the values of other fit statistics are then considered. The extent to which the minimum discrepancy value is acceptable is then determined and the other fit statistic values assessed. The line of fit statistic value consideration has been followed for every model, measurement and structural, included in this chapter.

The fit indices selected for evaluation of the structural model are similar to those used to determine the fit of the measurement model with the data from which it was estimated. The  $\chi^2$  statistic, the minimum discrepancy statistic (CMIN/df), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) have been included. The need to address the issue of parsimony in the structural model, however, has led to the inclusion of the Parsimony Goodness-of-Fit Index (PGIF) in the evaluation of the model hypothesised for the present study. To complement the PGFI measure, the Goodness-of-Fit Index (GFI) and Adjusted Goodness-of-Fit (AGFI) measures have also been included for evaluation of the structural model. The GFI, AGFI and PGFI are further discussed in the consideration of the structural model of the study later in this chapter.

## FIT OF THE MEASUREMENT MODEL

The purpose of this section is to consider the fit of the measurement models for the five latent variables (*Organisational Characteristics Model; Evaluator Characteristics Model; Stakeholder Characteristics Model; Evaluation Study Characteristics Model; and Use of Information Model*). The values of the fit indices considered in determining the fit of the five measurement models of the study are summarised in Table 7.1. The remaining latent variables of the

model (*Program Theory; Process Use; Commitment to Study (Pre); Commitment to Study (Post); Expected Study Involvement; Study Involvement; Influence of Study Findings; Influence of the Use of Program Theory in the Final Report; Influence of Involvement in Program Theory Elaboration*) are indicated by a single observed variable. The values of the latent variables with one indicator are the true scores for the variables, which we do not know. These are the scores that would be if there were no errors due to, for example, coding mistakes, misinterpretation of questions or intentional/strategic answers given which exaggerate or minimise (Hayduk, 1987).

**Table 7.1: Goodness of Fit Measures for Measurement Model Latent Variables**

Model	$\chi^2$	df	$p^1$	CMIN/df <sup>2</sup>	RMSEA <sup>3</sup>	TLI <sup>4</sup>
Organisational Learning Environment	6.79	6	0.19	1.13	0.09	0.96
Evaluator Characteristics	2.2	2	0.34	1.09	0.10	0.90
Stakeholder Characteristics	11.5	5	0.04	2.3	0.13	0.62
Evaluation Characteristics	2.2	2	0.34	1.07	0.05	0.98
Use Characteristics	1.37	1	0.24	1.37	0.10	0.98
Notes <sup>1</sup> A probability value for Chi Square test smaller than 0.05 indicates the null hypothesis that the model fits the data should be rejected. <sup>2</sup> Holmes-Smith (2001) suggests CMIN/df values should be greater than 1 but less than 2. However, he considers values between 2 and 3 are to be indicative of reasonable fit while values less than 1 indicate over fit. <sup>3</sup> The more generous upper threshold (0.1) of Browne and Cudeck (1989) has been used in the evaluation of model fit. <sup>4</sup> Generally, adjunct fit indexes (TLI) range between 0 and 1.0. A value of 0.90 is widely considered to be the threshold an index must exceed before a model is considered to be consistent with the observed data from which it was estimated (Hoyle, 1995; Byrne, 1995).						

## ORGANISATIONAL CHARACTERISTICS MODEL

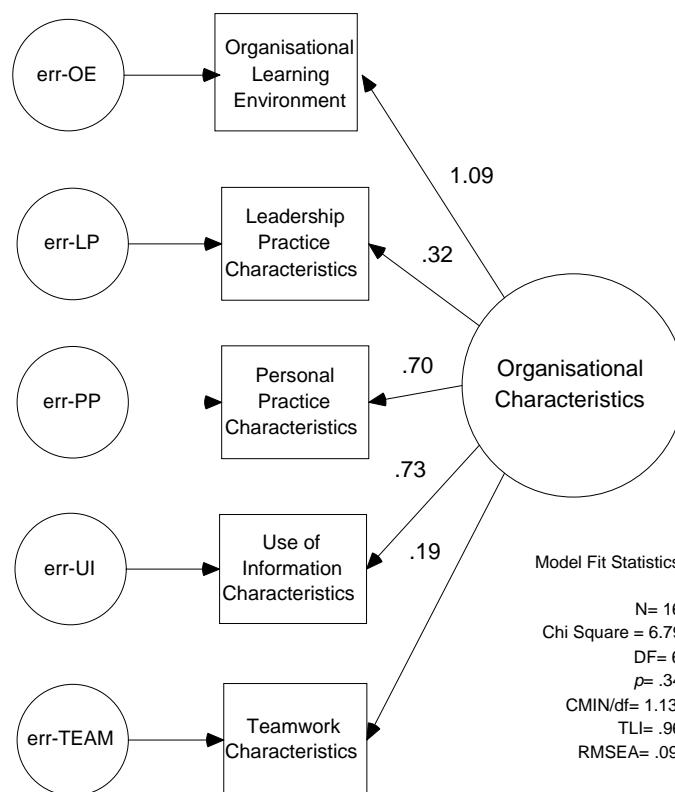
The *Organisational Characteristics* latent variable is predicted by the five scales of the Organisational Environment Characteristics Interview Schedule. Each scale provides the measure for one of the five observed variables

predicting the *Organisational Learning Characteristics* latent variable. The five scales focus on the Organisational Learning Environment characteristics, Leadership Practice characteristics, Personal Practice characteristics, characteristics of Use of Information, and characteristics of Teamwork. Of the five scales Organisational Learning Environment is the strongest predictor of the latent variable (regression weight 1.09). So, for instance, when the value of Organisational Learning Environment observed variable increases by 1, the value of the Organisational Characteristics latent variable increases by 1.09. The Personal Practice Characteristics and Use of Information Characteristics observed variables are moderate predictors of the Organisational Characteristics latent variable, while Leadership Practice Characteristics and Teamwork Characteristics are poorer predictors of the latent variable. The overall fit statistics information for the Model indicates the Model is consistent with the observed data from which it has been estimated. A reliability analysis of the five measures found all five to be correlated ( $\alpha = 0.77$ ). The RMSEA value is a little high, although the other fit indices indicate the Model is a good fit with the data. The diagram of the Organisational Characteristics Measurement Model with a summary of the Model fit statistics has been included in Figure 7.1. Note: a Sans Serif font has been applied in the text of this dissertation in order to discern the observed variables from the latent variables.

#### EVALUATOR CHARACTERISTICS MODEL

The *Evaluator Characteristics* latent factor is defined by four observed variables measured on scales included in the Evaluator Interview Schedule. The four scales focus on characteristics of the evaluator's practice (three sub-scales) and characteristics of the evaluator's perception of stakeholder involvement in the evaluation study. The model fit statistics indicate the model is consistent with the observed data from which it has been estimated. The Evaluator Practice Scale 1 is a very strong predictor of the *Evaluator Characteristics* latent variable, while the remaining three observed variables are very weak predictors. In fact, the Evaluator Practice Scale 2 has a slight negative regression weight value to the latent variable. A reliability analysis of

the four scales found them to be somewhat correlated ( $\alpha = 0.50$ ). Although the Evaluator Practice Scale 2 does not correlate with the Evaluator Practice Scale 1 or the Evaluator Practice Scale 2, it does correlate with the Perception of Stakeholder Involvement Scale and with Evaluator Practice Scale 1. Therefore, to maintain the breadth of the focus of the scale items all four scales were included as predictors of the *Evaluator Characteristics* latent variable. The *Evaluator Characteristics* Measurement Model with a summary of the Model fit statistics is represented in Figure 7.2.

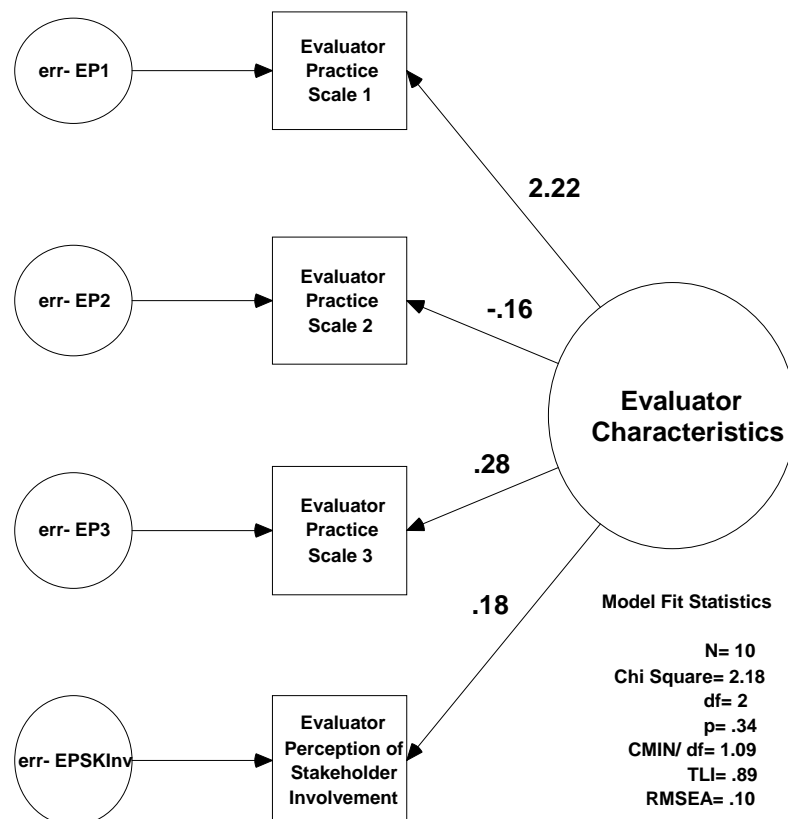


**Figure 7.1: Organisational Characteristics Measurement Model**

#### STAKEHOLDER CHARACTERISTICS MODEL

The *Stakeholder Characteristics* latent factor is defined by five scales included in the Stakeholder Interview Schedule 1 to gain information about the stakeholders of the evaluation studies included as cases in this study. The five

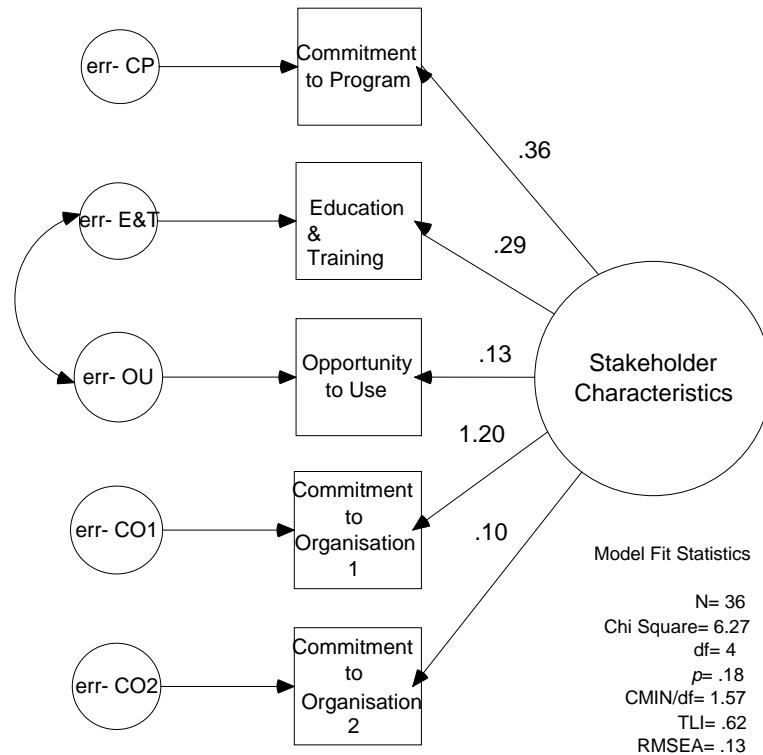
scales, each proving a measure for one of the five observed variables indicating the *Stakeholder Characteristics* latent variable, focus on stakeholders': commitment to the program, education and training, opportunity to use information, and commitment to the organisation (two sub-scales). The fit statistics of the first analysis of the *Stakeholder Characteristics* Model indicated the Model was not a good fit of the data. In particular, the  $p$  value is less than the 0.05 threshold above which the Model is considered acceptable. However, AMOS output indicated a cross-loading of the error variances of the Education and Training Scale and the Opportunity to Use (Information) Scale. In consideration of the two scales, it was determined there was likelihood that the conceptual focus of the scales are linked, as they are two sub-sections focusing on training and experience with research and evaluation in the interview schedule. To account for the cross-loading, the two error variance variables were co-varied in the second analysis of the model.



**Figure 7.2: Evaluator Characteristics Measurement Model**

The fit statistics resulting from the second analysis were more favourable as the fit statistics of the final Stakeholder Characteristics Model indicate it is a near fit. The  $p$  value is above the threshold suggested for model acceptance and the CMIN/df value is within the suggested range for Model acceptance. However, the TLI and RMSEA values are problematic. The TLI of 0.62 is less than the 0.90 threshold it should exceed before a Model is considered to be consistent with the data from which it was estimated. However, numerous authors recognise that the TLI frequently rejects models when sample sizes are very small (Bentler, 1990; Hu and Bentler, 1995), which is the case in this instance. Furthermore, the Bollen's incremental fit index (IFI) value (Bollen, 1989) for the Model (0.90) is very close to one, indicating the Model is a good fit of the data. Although Bollen's IFI is interpreted in the same way as the TLI using the same threshold values, Hoyle and Panther (1995) consider it less variable than the TLI in small samples and more consistent across estimators. An awareness of the small sample sizes of this study is important in interpretation of the Measurement Model fit statistics. Therefore, although the TLI value for the Model does not meet the criterion for accepting the Model as a good fit of the data, all other fit statistics used to evaluate the models of this study and other fit statistics, indicate the Model to be a good fit with the data. In addition, The RMSEA value is a little high, though the other fit indices indicate the Model is a good fit with the data.

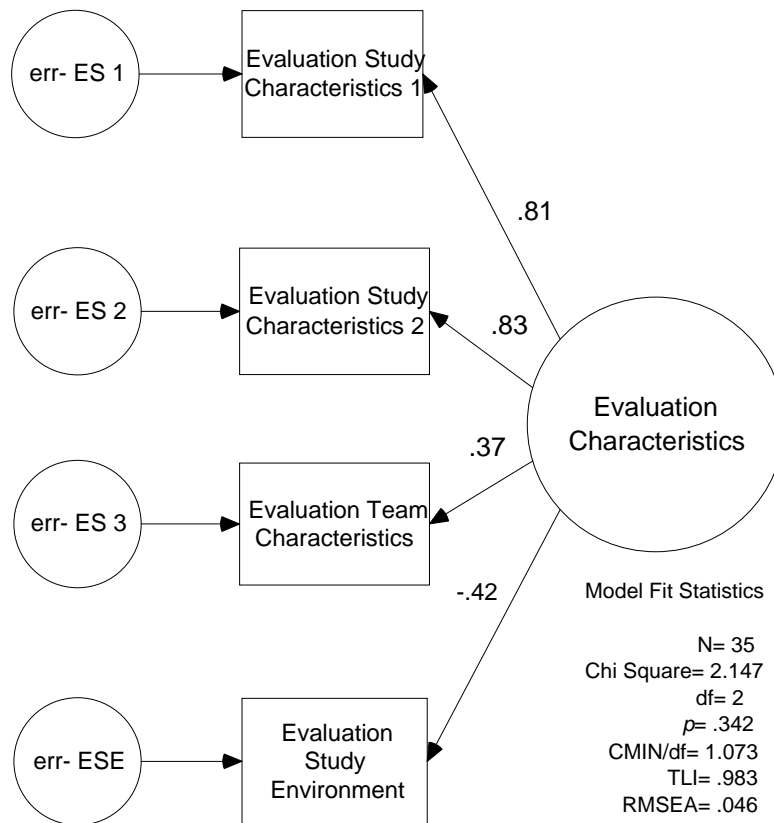
The regression weights indicate the Commitment to the Organisation 1 observed variable is a very strong predictor of the *Stakeholder Characteristics* latent variable. The other three observed variables are weak to moderate predictors. While a reliability analysis of the five observed variables found them not to be correlated ( $\alpha = 0.38$ ), a review of the inter-item correlations found the items correlated in such a way as to indicate that they measure separate yet related areas of focus. The Stakeholder Characteristics Measurement Model is represented in Figure 7.3 together with a summary of the Model fit statistics.



**Figure 7.3: Stakeholder Characteristics Measurement Model**

#### EVALUATION STUDY CHARACTERISTICS MODEL

The *Evaluation Study Characteristics* latent factor is indicated by four scales included in the Stakeholder Interview Schedule 2. Each of the scales provides a measure for an observed variable indicating the *Evaluation Study Characteristics* latent variable. Two of the four scales focus on the Study Characteristics, and one each on Evaluation Team Characteristics and the Evaluation Study Environment. The overall fit statistics information for the Model indicate the Model is consistent with the observed data from which it has been estimated. The *Evaluation Characteristics* Measurement Model with a summary of the Model fit statistics is represented in Figure 7.4.



**Figure 7.4: Evaluation Characteristics Measurement Model**

#### USE OF EVALUATION INFORMATION MODEL

In consideration of the Importance of Use and Likelihood of Use observed variables, a decision was made to combine the two to represent one latent variable labelled the *Use of Evaluation Information*. A reliability analysis of the two scales ( $r_{ii} = 0.83$ ) indicated the decision to combine the two to be a reasonable one.

In the reliability analysis of the scales presented in Chapter V, Item 78/79 was deleted from the Likelihood of Use scale due to low inter-item and item-total correlation values. This reduced the number of scale items from thirteen to twelve. For the purposes of dividing the scales into equal length sub-scale the corresponding item (Item 80/81) has been removed from the Importance of Use scale. The two scales were each divided into two sub-scales, each containing six items.



Likelihood of Use scale was divided in consideration of each item's item-total correlation value in the scale comprised of all items. The items were paired, beginning with the item with the lowest value. In the earlier analysis of the Likelihood of Use scale and the Importance of Use scale (refer to Chapter V), two items were combined to determine a Likelihood of Use value and an Importance of Use value for each of the thirteen manners of use that were the focus of the scales. The item combinations (e.g. 82/83) represent a single value.

The first sub-scale of *Likelihood of Use (L1)* is comprised of the values of item combinations 82/83, 86/87, 94/95, 102/103, 110/111, 122/123. In a reliability analysis of the sub-scale, Cronbach's Alpha ( $\alpha$ ) was found to be 0.75. All items correlated with three or more items ( $r_{ii'} > 0.30$ ) with the exception of Item 122/123, which correlated with only one other item ( $r_{ii'} > 0.30$ ). Item total correlations for all items were good ( $r_{it} > 0.30$ ), again with the exception of Item 122/123 ( $r_{it} = 0.28$ ).

The second sub-scale of *Likelihood of Use (L2)* is comprised of the values of item combinations 90/91, 98/99, 106/107, 114/115, 118/119, 126/127. In a reliability analysis of the sub-scale, Cronbach's Alpha ( $\alpha$ ) was found to be 0.72. All items correlated with two or more items ( $r_{ii'} > 0.30$ ) with the exception of Item 126/127, which correlated with only one other item ( $r_{ii'} > 0.30$ ). All item total correlations were good ( $r_{it} > 0.30$ ).

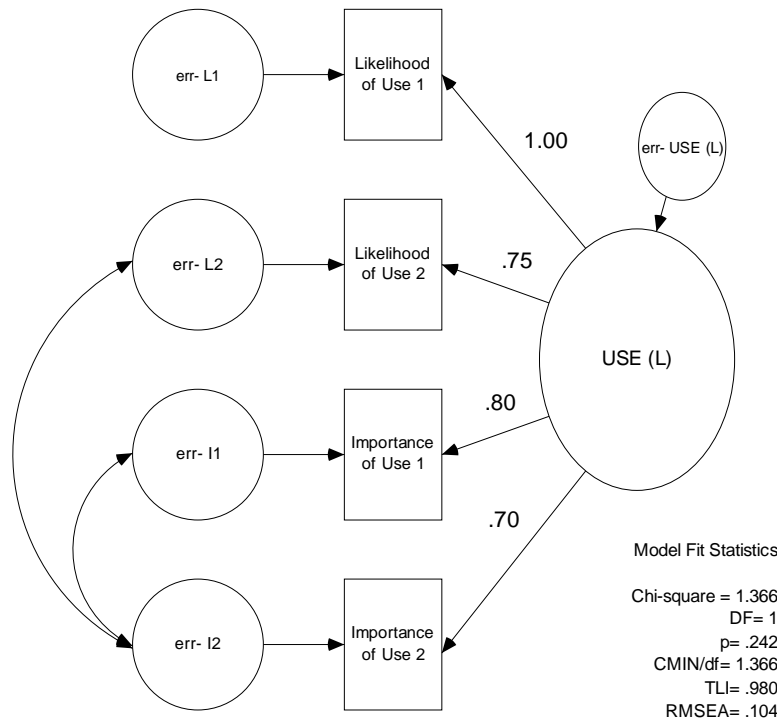
The first sub-scale of *Importance of Use (I1)* is comprised of values 84/85, 88/89, 96/97, 104/105, 112/113, and 124/125. In a reliability analysis of the sub-scale Cronbach's Alpha ( $\alpha$ ) was found to be 0.77. All items correlated with two or more items ( $r_{ii'} > 0.30$ ). All item total correlations were good ( $r_{it} > 0.29$ ).

The second sub-scale of *Importance of Use (I2)* is comprised of values 92/93, 100/101, 108/109, 116/117, 120/121, and 128/129. In a reliability analysis of the sub-scale Cronbach's Alpha ( $\alpha$ ) was found to be 0.77. All items correlated

with two or more items ( $r_{ii'} > 0.30$ ). All item total correlations were good ( $r_{it} > 0.29$ ).

In the first analysis of the Model, the fit statistics all indicated the Model to be a poor fit. A review of the Modification Indices indicated the fit of the Model would be improved if err- L2 and err- I2 were co-varied (*MI 11.52, Expected Par Change 11.71*). Conceptually, this relationship is reasonable as the items of the two scales were taken from the same clusters. As mentioned above, a cluster of four items focused on each of the thirteen types of use and were included in the original scale.

The Model was run a second time, with err- L2 and err- I2 co-varied. Again, the fit statistics indicated the Model to be a poor fit with the data from which it was estimated. A review of the modification indices indicated the fit of the Model would be improved if err- U1 and err- U2 were co-varied. Conceptually, co-varying the error terms of these two observed variables made sense as both are concerned with the measurement of the importance of use. The final fit statistics information for the Model indicate the re-specified Model to be a good fit with the observed data from which it has been estimated. A review of the modification indices indicated no further re-specification of the Model would improve its fit. The Likelihood of Use Scale 1 is the strongest predictor of the *Use of Information* latent variable, however, the remaining three scales are also very strong predictors. A reliability analysis of the four scales found them to be highly correlated ( $\alpha = 0.91$ ). The I/L Use Measurement Model is represented in Figure 7.5 with a summary of the Model fit statistics.



**Figure 7.5: Importance/ Likelihood of Use Measurement Model**

## SUMMARY

The five Measurement Models developed for the study are: the Organisation Characteristics Model, the Evaluator Characteristics Model, the Stakeholder Characteristics Model, the Evaluation Study Characteristics Model, and the Use Model. The fit statistics with regards to the Models indicate all five are consistent with the data from which they are estimated. The exceptions are the TLI value for the Stakeholder Characteristics Model, and the RMSEA values for the Organisational Characteristics Model and the Stakeholder Characteristics Model. In consideration of the other fit information indicating these Models are a good fit, decisions were made to accept the Models.

## TESTING OF THE STRUCTURAL MODEL

The diagrams of the Structural Models included in this dissertation contain a large amount of detail due to the advanced graphics package that distinguishes AMOS from other software programs used to undertake structural equation modelling. The benefit to the reader is the wealth of detail regarding the model and its statistics included in the text. However, it is necessary to identify the information included in the diagrams in this chapter to aid understanding of the diagrams.

The large ovals in the diagram represent the latent variables of the model. The squares linked to each by one-way arrows represent the observed or measured variables. Observed variables are usually represented by scale, sub-scale or item scores. The one-way arrows designate structural regression coefficients and are displayed adjacent to the one-way arrows. These coefficient values represent the amount of change expected in the observed variable for change in the predictor latent variable, indicating a possible causal link. In some instances the coefficients have been set to a value of 1 rather than being left free for AMOS to estimate in the analysis of the structural equation. This has been done when there is only one latent predictor variable for the observed variable and is a strategy recommended by a number of researchers using AMOS for structural equation modelling (Bryne, 2001; Holmes-Smith, 2001). In such instances, the single observed variable is the sole measure of the latent variable. The small circle linked to each observed variable represents the measurement error (random and non-random) on the observed variables.

Latent variables *causing* change in another latent variable are exogenous, whereas the affected latent variables are endogenous. Endogenous latent variables have a small circle linked to them representing the *residual error term* (or unexplained variance, the complement of explained variance accounted for by the model) of the exogenous variable in predicting the endogenous variable. Curved bidirectional arrows linking variables indicate co-variance between the linked variables, error terms, etc (Bryne, 2001).

Fit statistics employed in the evaluation of each model have also been included in the diagrams in the upper left corner. The fit statistics for the evaluation of the measurement models include the Chi Square and the probability of the Chi Square value ( $\chi^2$ ), the minimum discrepancy (CMIN/df), the Tucker-Lewis Index (TLI) and the root mean square error of approximation (RMSEA). For the evaluation of the more complex structural models, the fit statistics used for the evaluation of the measurement models have been included, and three additional fit statistics added. They are the Goodness-of-Fit Index (GFI), the Adjusted Goodness-of-Fit Index (AGFI), and the Parsimonious Goodness-of-Fit Index (PGFI).

The structural model is the second stage, or higher order, undertaking of the structural equation modelling process. Once the measurement model linking the relationships between the unobserved and observed variables has been defined, it is possible to undertake modelling specifying the relations between the unobserved variables. The structural model specifically focuses on the manner of direct or indirect influence particular latent variables cause in the values of specific other latent variables (Byrne, 2001; Cummings, 1997; Holmes-Smith, 2001).

#### **FIT INDICES EMPLOYED FOR STRUCTURAL MODEL EVALUATION**

Generally, structural equation models are evaluated using statistical tests in three ways. First, overall fit statistics are employed to determine the adequacy of a model. The fit indices selected for evaluation of the structural model are similar to those used to determine the fit of the measurement model with the data from which it was estimated. The Chi Square statistic, the minimum discrepancy statistic (CMIN/df, acceptable threshold value between 1 and 3), the Tucker-Lewis index (TLI, acceptable threshold value between 0 and 1), and the root mean square error of approximation (RMSEA, acceptable threshold value less than 0.10) have been included. The discussion of each fit index, the threshold values for determination of model fit, and a summary of

the literature relevant to each, has been included in the above section focusing on the measurement model. The threshold values for each fit statistic adopted for the evaluation of the measurement model are the same for the evaluation of the structural model, and have been included as footnotes in the tables representing the model fit.

In addition, overall fit statistics and rules of parsimony are used to compare alternative structural models to each other. Parsimony is determined in terms of the number of parameters that must be estimated. As a rule, a variety of statistics should indicate statistically-significant components and good overall model fit with the most parsimonious model being favoured. Generally, first order, or measurement models tend to be more parsimonious, as fewer parameters are required to be estimated. Furthermore, with regards to the measurement model, all the items from the scales contributing the measured, observed factors have been developed to measure the same trait, and are therefore highly correlated. However, structural models considering the relations between unobserved variables often require many parameters to be estimated (Hull, Tedlie and Lehn, 1995). Finally, the individual scales are associated with dimensions of a higher order construct. Although the scales share a certain amount of variance with regards to the general higher order construct, they retain a reasonable amount of unique variance.

Therefore, the need to address the issue of parsimony in the structural model has led to the inclusion of the Parsimony Goodness-of-Fit index (PGFI) in the evaluation of the model hypothesised for the present study. To complement the PGFI measure, the Goodness-of-Fit Index (GFI) and Adjusted Goodness-of-Fit Index (AGFI) measures have also been included for evaluation of the structural model. The GFI, AGFI and PGFI are further discussed in the following section.

Finally, the second way that structural models are evaluated using fit statistics is in the judgment of individual aspects of a particular model. These fit statistics include the *modification indices* and *residual co-variances* referred to

in consideration of modifications to the model, to increase its fit with the data from which it has been estimated (Bryne, 2001; Information Technology Services at the University of Texas at Austin, 2002; Holmes-Smith, 2001). The implications of these measures are reported in the assessment of each model, as appropriate. If modification indices and residual co-variance values are within normal limits, no modification of the model is likely to be undertaken.

#### GOODNESS-OF-FIT (GFI) AND ADJUSTED GOODNESS-OF-FIT (AGFI)

GFI and AGFI are considered to be absolute fit indices as they compare the proposed model to no model at all and measure the relative amount of variance and co-variance jointly explained by the model. The GFI was proposed by Jöreskog and Sörbom (1984) as an absolute fit index for Maximum Likelihood (ML) estimation, though Generalised Least Squares (GLS) versions of the GFI and AGFI have been presented by Tanaka and Huba (1985). Marsh et al. (1988) found the GFI performed better than any other absolute fit index they studied. In addition, the GFI provides an *intuitive* interpretation very like the  $R^2$  value associated with multiple regression models (Tanaka and Huba, 1985). Furthermore, Hu and Bentler (1995) and Tanaka (1987) found the GFI performed consistently across ML and GLS estimation methods at all sample sizes when both common and unique variates were distributed independent of each other. However, it has also been reported that when the latent variables are dependent, GFI behaved inconsistently with sample sizes of 250 or less (Hu and Bentler, 1995).

The AGFI is very similar to the GFI, though an adjustment has been made to take into account the degrees of freedom of the model. The AGFI also incorporates a penalty function to account for the inclusion of an additional parameter. The AGFI is reported to perform consistently with the ML estimation method when the latent variables are independent and the sample sizes are above 250 (Hu and Bentler, 1995).

The maximum value of the GFI and AGFI is 1 and though typical values are greater than 0, it is possible for them to be less than zero. Holmes-Smith (2001) suggests values between 0.90 and 0.95 may also indicate satisfactory fit, though Hu and Bentler (1995) found the adoption of a cut-off value of 0.90 resulted in over rejection of true models when latent variables were dependent. Values close to 1 are considered to be indicative of a good fit. Some have cautioned that both GFI and AGFI values can be overly influenced by sample size (Hu and Bentler, 1995). In particular, Hu and Bentler (1995) found the AGFI and GFI are less reliable when sample size is small.

#### PARSIMONY GOODNESS-OF-FIT INDEX (PGFI)

James, Mulaik and Brett (1982) introduced the Parsimony Goodness-of-Fit Index (PGFI) measure in response to the issue of parsimony in SEM. The value of the PGFI is the extent to which the index takes into account the number of estimated parameters in the hypothesised model in terms of the overall model fit. The PGFI value represents the goodness of the model fit measured by the GFI and the parsimony of the model. Numerous authors suggest complex models be penalized by multiplying the fit index by a parsimony ratio specified as the degrees of freedom of the target model, relative to the total number of degrees of freedom in the data (e.g. James et al., 1982, Williams and Holahan, 1994). Parsimony-based index values are often lower than the acceptable threshold for other normed indices of fit. For instance, goodness of fit indices with a value of around 0.90 range can be linked with parsimonious fit indices around 0.50 (Bryne, 2001; Mulaik et al., 1989). In fact, as James et al. (1982) note, a just identified model with zero degrees of freedom would have a PGFI value of 0.00, even when it has a normed index value of 1. Therefore, a threshold value of 0.50 has been accepted for this study.



## INFORMATION CONSIDERED IN INTERPRETATION OF FINDINGS

In the consideration of model fit, priority has been given to certain fit statistics. The two fit statistics, however, considered first in model assessment are the  $p$  value associated with the Chi Square and the degrees of freedom. These two values give a good snapshot of the overall model fit (Bryne, 2001). This is the information provided in the AMOS output model summary. The extent to which the minimum discrepancy value is acceptable is then determined and the other fit statistic values assessed. The line of fit statistic value consideration has been followed for every model included in this study.

With regards to the structural models, further information is considered with a view to interpretation of findings. Specifically, effect values and path regression values yield important information. The thresholds considered in the interpretation of these statistics are considered below.

### INDIRECT, DIRECT AND TOTAL STANDARDISED EFFECT OF LATENT PREDICTOR VARIABLES ON THE OUTCOME VARIABLE

Tables representing the standardised indirect, direct and total effect size of each latent predictor variable for each version of the Model have been included below. The indirect value represents the size of the effect of the latent predictor variable on the outcome variable through other predictor variables. The direct value represents the extent to which the predictor variable has sole influence on the outcome variable. Finally, the total effect value reflects the influence of the predictor variable on the outcome variable, summing both the indirect and direct effect values.

There is some ambiguity with regards to the suggested threshold values for determining the significance of effect size (Hinkle et al., 1988). Cohen (1965), defining effect size as the degree to which a phenomenon exists, suggests an effect of 0.25 or more be considered small, an effect of 0.50 or more as medium and an effect of 1.00 or more as large. Alternatively, Holmes-Smith (personal communication, July 24, 2007) suggests any value below 0.20

should be considered weak, a value between 0.20 and 0.30 is considered small, and while values between 0.30 and 0.50 indicate a stronger effect, values between 0.50 and 0.80 indicate a very strong effect.

With regards to the present study, the threshold values of Holmes-Smith (personal communication, July 24, 2007) have been adopted for the assessment of effect sizes for three primary reasons. First, this study relies on a small sample size which may have the influence of yielding less specific statistical values. Therefore, the adoption of less stringent thresholds for assessment widens the opportunity for the identification of possible areas having value for further research. If a lower value were adopted as the significance threshold level, it is likely that some areas for further research with value might not be identified. Second, the findings of this study rest on the investigation of only three case studies and therefore the degrees of freedom available for the statistical analysis of the data are small. Third, although there is considerable theoretical literature available with regards to the influence of the use of program theory in the undertaking of an evaluation, there is almost no empirical research. With further research, the adoption of less generous significant levels might be appropriate, but at this early stage, the intended goal of the research is to identify areas with value for further research.

#### REGRESSION WEIGHTS

Standardised regression weights indicating the predictive power of one latent variable on another, are displayed on the paths of the diagram. For example, a standardised regression weight value of 0.40 indicates the value of the predicted variable increases by 0.40 standard deviations when the predictor variable value goes up by one standard deviation. The probability of attaining the critical ratio value (estimation value divided by standard error) determines the significance of regression weight value. For example, if the probability of getting a critical ratio as large as 1.44 in absolute value is 0.15, the predictive power of the variable is not significantly different from 0 at the 0.10 level (two-tailed).

With regards to regression weights, the 0.10 significance level threshold has been adopted here rather than the more common and rigorous 0.05 level for the same three reasons given in the section above, focusing on effect size. First, the small sample size of the study may compromise the accuracy of the statistics the study yields. Second, the study rests on the investigation of only three case studies and the degrees of freedom available for the statistical analysis of the data are small. Finally, this study deals with a relatively unexplored area of research. One expected outcome of the present study is to identify areas with value for further research.

### **VARIABLES OF THE MODEL**

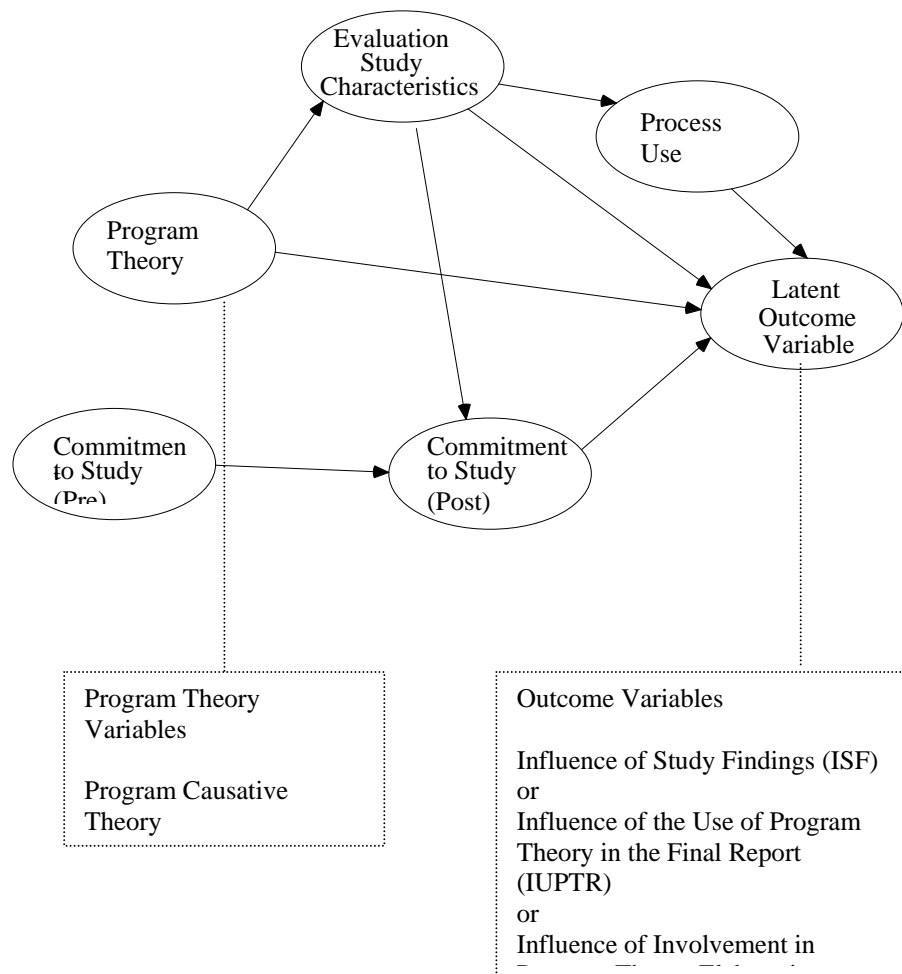
The latent predictor variables of the Core Model are *Program Theory*, *Evaluation Study (Characteristics)*, *Process Use*, *Commitment to the Study (Pre)* and *Commitment to the Study (Post)*. The Model could have included more latent predictor variables. Certainly, more have been identified as having an influence on an evaluation study and its outcomes (e.g. Cousins and Leithwood, 1986, 1993; Hudson-Mabbs, 1993). However, due to the conditions for the undertaking of the structural equation modelling process, the Model for testing could not be too complicated. It was important that the Model be parsimonious with a view to the most efficient use of the information gathered. First, the degrees of freedom for the undertaking of the statistical calculations for the SEM would not allow for the inclusion of a large number of latent variables in the Model. Second, the small sample size upon which the study is based has a potential to compromise the accuracy of the SEM statistics. Third, the study rests on the investigation of only three case studies. Finally, this study deals with a relatively unexplored area of research. One expected outcome of the present study is to identify areas with value for further research. A generic version of the Core Model of the study is presented in Figure 7.6. The Influence of each variable in the Model is discussed below.

## PROGRAM THEORY

The use of program theory in an evaluation study, both causative theory and implementation theory, is expected to have a direct influence on the characteristics of the evaluation study and the outcomes of the study. In particular, the use of program theory is expected to influence stakeholder perception of the relevance and quality of the evaluation by encouraging stakeholder understanding of, and insight into, the program activities and logic, and by focusing the information collected in the evaluation process to that indicated to be most relevant by the program theory. Furthermore, by linking the findings of the study back to the program theory, it is expected that an increased use of program theory in the evaluation study will lead to a greater understanding of the implications for the use of the evaluation information, thereby increasing the influence of the information.

## COMMITMENT TO THE STUDY (PRE & POST)

*Commitment to the Study (Pre)* is expected to have a direct influence on the stakeholders' commitment to the study post-conduct, and an indirect influence on the outcomes of the study. Specifically, the higher a stakeholder's level of commitment to the evaluation study at the beginning of the study, the more likely they are to be committed at a higher level to the evaluation study following its conduct. Finally, the higher the level of stakeholder *Commitment to the Study (Post)*, following its conduct, the more likely they are to be influenced by an evaluation's information, or overall outcome. *Stakeholder Commitment to the Study (Post)* is also directly influenced, in the Core Model, by the *Evaluation Study Characteristics*. This link is discussed further in the *Evaluation Study Characteristics* section below.



**Figure 7.6 Generic Version of Core Structural Equation Model**

#### EVALUATION STUDY CHARACTERISTICS

The characteristics of an evaluation study are expected to have influence an on the *Process Use* of the study, *Commitment to the Study (Post)* conduct, and the outcomes of the study. For instance, if stakeholders perceive an evaluation study to be of sound quality, the information of the study to be relevant to the needs of the program, the environment of the evaluation to be open and transparent, and the practice of the evaluation team credible, then it is more likely those involved in the evaluation process will put what they have learned from their involvement to use. Similarly, if stakeholders have positive perceptions of the evaluation study characteristics, they are more likely to be committed to the study. Furthermore, by having a direct positive influence on *Process Use* and *Commitment to the Study (Post)*, there is an avenue for the

*Evaluation Study Characteristics* variable to have an indirect, as well as direct, influence on the outcomes of an evaluation study.

#### PROCESS USE

Simply, stakeholders influenced by their involvement in an evaluation study are more likely to be influenced by the outcomes of the study. With this involvement comes an understanding of the evaluation processes and insight. This knowledge is more likely to facilitate the influence of an evaluation study's information on the stakeholder, following the dissemination of the final report.

#### OUTCOME VARIABLES OF THE CORE MODEL

Four different areas of use or influence have been included in the model. They include the *Influence of Study Findings (ISF)*, *Influence of Use of Program Theory in the Final Report (IUPTR)*, the *Influence of Involvement in Program Theory Elaboration (IIPTE)*, and the *Use of Evaluation Information (UI)*. Each outcome measure, with the exception of the *UI* outcome variable, is represented in the model by a single observed variable, composed of one scale. The *UI* variable is represented by four observed variables. The development and testing of the four outcome variables is described in Chapter V.

The decision to develop four parallel models for testing, each with a different outcome measure, was taken because the small sample size of this study made it necessary to reduce the complexity of the models tested. Each outcome measure is tested independently rather than including all four in the model simultaneously. In addition, had only one structural model incorporating four outcome variables been developed, valuable information might have been lost. Although a correlation of the four outcome variables found them to be highly correlated ( $r = 0.71$ ), maintaining a separateness of the outcome variables in the Model versions is expected to enable insights that might otherwise have been lost in the complexity. Each latent outcome variable is considered below.

#### MODEL VERSION 1: INFLUENCE OF STUDY FINDINGS (ISF)

The first core model developed for the study includes the latent outcome variable *Influence of Study Findings (ISF)*. This outcome variable is intended to focus on the more ‘traditional’ forms of evaluation influence detailed earlier, i.e. conceptual, instrumental and strategic (Alkin, 1985; Cousins and Leithwood, 1986, 1993; Cummings, 1997; Hudson-Mabbs, 1993; Leviton and Hughes, 1981; Patton, 1997).

#### MODEL VERSION 2: INFLUENCE OF USE OF PROGRAM THEORY IN THE FINAL REPORT (IUPTR)

This variable is a primary focus of this study. This outcome variable is expected to focus on the influence of the use of program theory included in the final evaluation report, on the thinking, practice and decisions of the stakeholders.

#### MODEL VERSION 3: INFLUENCE OF INVOLVEMENT IN PROGRAM THEORY ELABORATION (IIPTE)

The influence of stakeholder involvement in any processes to elaborate or detail the theory underpinning the program, is expected to have an influence on the use of the evaluation information. The literature suggests stakeholder involvement in any process to elaborate the program theory or achieve consensus regarding program theory, may increase their awareness and understanding of the program, encourage reflection regarding their practice, lead to changes in their practice and inform decisions regarding the program evaluated (Cummings et al., 2001; Huebner, 2000; Milne, 1993).

MODEL VERSION 4:   USE OF EVALUATION INFORMATION (UI)  
IMPORTANCE OF USE (IU) AND THE LIKELIHOOD OF  
USE (LU)

This latent outcome variable is a combination of two observed variables: *Importance of Use* and *Likelihood of Use*, as utilised by Cummings (1997). The items of the scales focus on the importance and likelihood of a variety of possible uses of an evaluation's information.

### SUMMARY

Eight versions of the Core Model have been detailed in this section. Each outcome variable will be included in two versions of the Core Model, one including *Implementation Theory* as a *Program Theory* predictor variable, and the other including *Causative Theory*. This strategy is expected to facilitate model fit, in view of the small sample size of the study. Furthermore, in consideration of the simpler models, it is expected that the influences of the predictor variables on the outcome variables, both direct and indirect, will be clearer. The eight versions of the Core Model are summarised in Table 7.2.

**Table 7.2: Eight Versions of the Core Model**

Versions Number	Outcome Variable	Latent Program Theory Predictor Variable
1a	Influence of Study Findings	Causative Program Theory
1b	Influence of Study Findings	Implementation Program Theory
2a	Influence of Use of Program Theory in Final Report	Causative Program Theory
2b	Influence of Use of Program Theory in Final Report	Implementation Program Theory
3a	Influence of Involvement in Program Theory Elaboration	Causative Program Theory
3b	Influence of Involvement in Program Theory Elaboration	Implementation Program Theory
4a	Use of Information	Causative Program Theory
4b	Use of Information	Implementation Program Theory



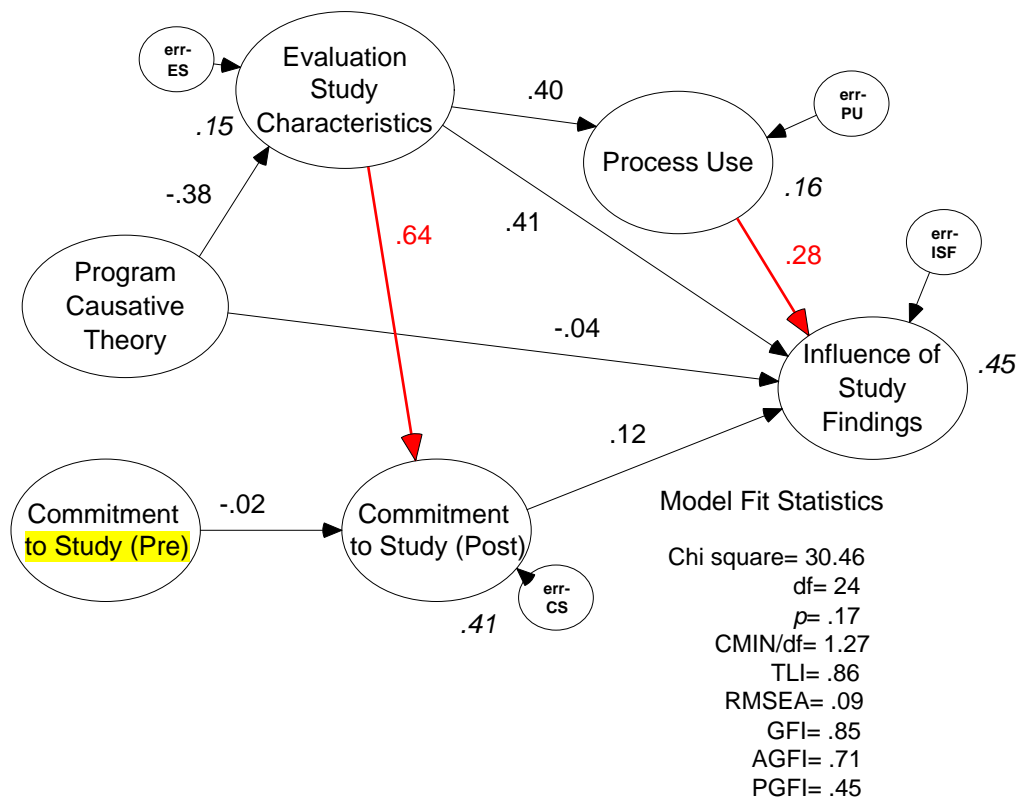
## TESTING OF THE EIGHT VERSIONS OF THE CORE MODEL

### CORE MODELS 1A & 1B

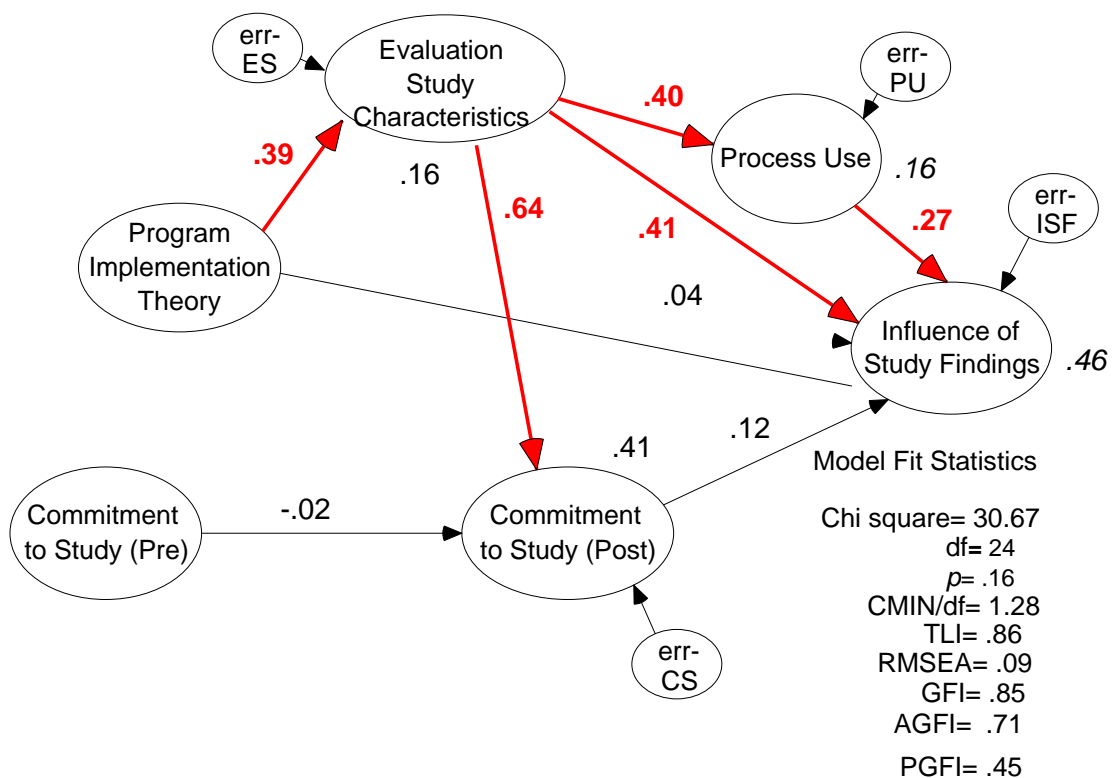
### INFLUENCE OF STUDY FINDINGS

In versions 1a and 1b of the Core Model, *Influence of Study Findings* is the outcome variable and *Program Causative Theory* is the predictive variable in Model 1a, and *Program Implementation Theory* as a predictive variable in Model 1b. The overall fit statistic information for Structural Model 1a and Structural Model 1b ( Figures 7.7 and 7.8) indicates the Models are generally consistent with the observed data from which they have been estimated. The *p*-value of both Models is greater than 0.05, the suggested lower threshold for model acceptance. The CMIN/df discrepancy value for both models is midway between 1 and 3, the suggested upper and lower ranges for the fit statistics. The RMSEA is below 0.10, the upper threshold above which it is suggested that the model be rejected. The TLI is very near the threshold for model acceptance (0.09). Finally, the AGFI, GFI and PGFI values are slightly low; though considering the size of the sample this is not unexpected.

The TLI, GFI, AGFI and PGFI values are just below the commonly suggested lower thresholds for model acceptance for each (TLI> 0.90, GFI> 0.90, AGFI> 0.90, and PGFI> 0.90) though much lower PGFI values are acceptable if other fit statistic values are good. The small sample size of the study is a factor considered in the assessment of the fit statistic values for the structural models, however. Furthermore, the RMSEA values are slightly higher than the commonly accepted upper threshold value (RMSEA<0.05). Structural Models 1a and 1b (representing the first two versions of the Core Model) with fit statistics and standardised regression weights are represented in Figures 7.7 and 7.8, respectively. The regression path with estimates indicated to be significant at the 0.10 level, have been marked in red on the structural model diagrams. Table 7.3 presents the regression path estimates of Structural Models 1a and 1b and the probability value of attaining each estimate by chance. Direct, indirect and total effects of latent predictor factors on the outcome variable are presented in Table 7.4.



**Figure 7.7: Structural Model 1a**



**Figure 7.8: Structural Model 1b**

**Table 7.3 Standardised Regression Weights for Model Version 1**

Regression Path Variables		Model 1a Estimate (Probability)	Model 1b Estimate (Probability)
Predictor Variable	Predicted Variable		
Program Causative Theory	Evaluation Study Characteristics	-0.38 (0.15)	
Program Implementation Theory	Evaluation Study Characteristics		0.39 (0.02)
Program Causative Theory	Influence of Study Findings	-0.04 (0.77)	
Program Implementation Theory	Influence of Study Findings		0.04 (0.77)
Evaluation Study Characteristics	Process Use	0.40 (0.15)	0.40 (0.02)
Evaluation Study Characteristics	Commitment to Study (Post)	0.64 (0.10)	0.64 ( $< 0.001$ )
Evaluation Study Characteristics	Influence of Study Findings	0.41 (0.16)	0.41 (0.06)
Process Use	Influence of Study Findings	0.28 (0.06)	0.27 (0.06)
Commitment to Study (Pre)	Commitment to Study (Post)	-0.02 (0.90)	-0.02 (0.90)
Commitment to Study (Post)	Influence of Study Findings	0.12 (0.52)	0.12 (0.53)

**Table 7.4 Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Study Findings (Model Version 1)**

<b>Theory Predictor :</b>	<b>Causative Theory (Version 1a)</b>			<b>Implementation Theory (Version 1b)</b>		
<i>Squared Multiple Correlation for Structural Equation</i>	<b>R<sup>2</sup> = 0.45</b>			<b>R<sup>2</sup> = 0.46</b>		
	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>
<b>Predictor Variable</b>						
<b>Program Causative Theory</b>	-0.04	-0.23	-0.27			
<b>Program Implementation Theory</b>				0.04	0.24	0.28
<b>Evaluation Study Characteristics</b>	0.41	0.19	0.60	0.42	0.19	0.60
<b>Process Use</b>	0.28	0.00	0.28	0.27	0.00	0.27
<b>Commitment to Study (Pre)</b>	0.00	0.00	0.00	0.00	0.00	0.00
<b>Commitment to Study (Post)</b>	0.13	0.00	0.13	0.12	0.00	0.12
<p>Note: Holmes-Smith (personal communication, July 24, 2007) suggests any effect size value below 0.2 be considered weak, a value between 0.2 and 0.3 be considered small, while values between 0.3 and 0.5 indicate a stronger effect values between 0.5 and 0.8 indicate a very strong effect.</p> <p>Cronbach's Alpha of the scale measuring the outcome variable <i>Influence of the Study Findings</i> indicates the scale reliability is substantial (<math>\alpha = 0.76</math>).</p>						

Core Model versions 2a and 2b replace *ISF* with *Influence of Use of Program Theory in Final Report* as the outcome variable. These versions do not fit the data as well as models 1a and 1b. The fit statistic values of the models, however, are close to meeting the suggested lower thresholds of the fit statistics. Therefore, the models and the data from them have been included here. In model versions 2a and 2b *Influence of Use of Program Theory in Final Report (IUPTR)* is the outcome variable predicted by the same five latent variables identified in the previous model. Structural Model 2a includes *Causative Program Theory* as a predictive variable, while Structural Model 2b includes *Implementation Program Theory* as a predictive variable.

The fit statistics values for version 2a generally indicate the Model is consistent with the observed data from which it has been estimated. Structural Models 2a and 2b, with fit statistics and standardised regression weights, are represented in Figures 7.9 and 7.10 respectively. Table 7.5 presents the regression path estimates and probability values of attaining each by chance for Structural Models 2a and 2b. The regression paths with estimates indicated to be significant at the 0.10 level have been marked in red on the structural model diagrams. Further insight into the influence of the predictor variables is given in the standardised direct effects, standardised indirect effects and standardised total effects values reported in Table 7.6.

Chi square= 33.10  
df= 24  
p= .10  
CMIN/df= 1.38  
TLI= .79  
RMSEA= .11  
GFI= .84  
AGFI= .69  
PGFI= .45

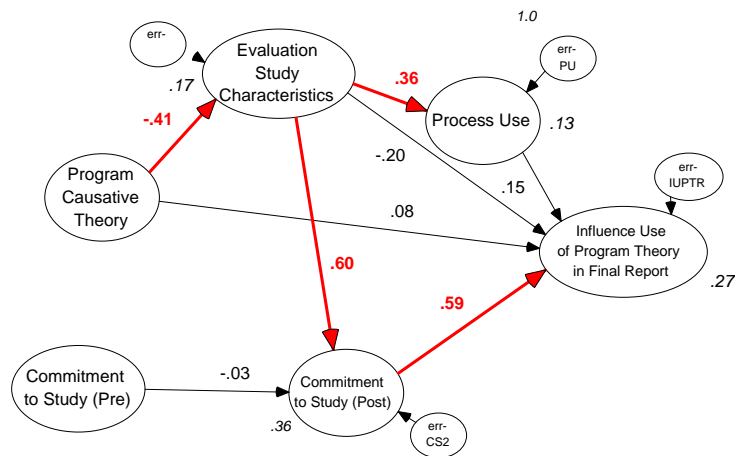


Figure 7.9: Structural Model 2a

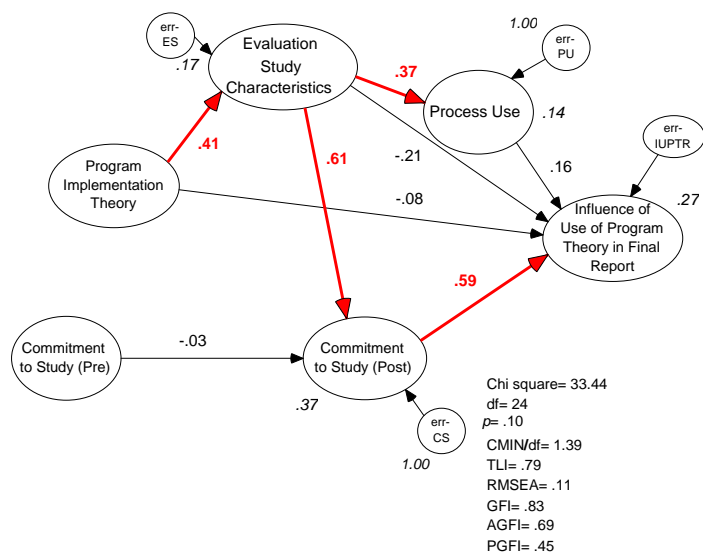


Figure 7.10: Structural Model 2b

**Table 7.5    Standardised Regression Weights for Model Version 2**

<b>Regression Path Variables</b>		<b>Model 2a Estimate (Probability)</b>	<b>Model 2b Estimate (Probability)</b>
<b>Predictor Variable</b>	<b>Predicted Variable</b>		
Program Causative Theory	Evaluation Study Characteristics	-0.41 (0.03)	
Program Implementation Theory	Evaluation Study Characteristics		0.41 (0.02) <sup>2</sup>
Program Causative Theory	Influence of Use of PT in the Final Report	0.08 (0.61)	
Program Implementation Theory	Influence of Use of PT in the Final Report		-0.08 (0.63)
Evaluation Study Characteristics	Process Use	0.36 (0.03)	0.37 (0.03)
Evaluation Study Characteristics	Commitment to Study (Post)	0.60 ( $< 0.001$ )	0.61 ( $< 0.001$ )
Evaluation Study Characteristics	Influence of Use of PT in the Final Report	-0.20 (0.28)	-0.21 (0.29)
Process Use	Influence of Use of PT in the Final Report	0.15 (0.32)	0.16 (0.32)
Commitment to Study (Pre)	Commitment to Study (Post)	-0.03 (0.82)	-0.03 (0.83)
Commitment to Study (Post)	Influence of Use of PT in the Final Report	0.59 (0.01)	0.59 (0.01)

**Table 7.6 Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Use of Program Theory in Final Report (Model Version 2)**

<b>Theory Predictor:</b>	<b>Causative Theory (Version 2a)</b>			<b>Implementation Theory (Version 2b)</b>		
<i>Squared Multiple Correlation for Structural Equation</i>	<b>R<sup>2</sup> = 0.27</b>			<b>R<sup>2</sup> = 0.28</b>		
	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>
<b>Predictor Variable</b>						
<b>Program Causative Theory</b>	0.08	-0.08	0.00			
<b>Program Implementation Theory</b>				-0.08	0.09	0.01
<b>Evaluation Study Characteristics</b>	-0.2	0.41	0.20	-0.21	0.42	0.21
<b>Process Use</b>	0.15	0.00	0.15	0.16	0.00	0.16
<b>Commitment to Study (Pre)</b>	0.00	-0.02	-0.02	0.00	-0.02	-0.02
<b>Commitment to Study (Post)</b>	0.59	0.00	0.59	0.59	0.00	0.59
<p>Note: Holmes-Smith (personal communication, July 24, 2007) suggests any effect size value below 0.2 be considered weak, a value between 0.2 and 0.3 be considered small, while values between 0.3 and 0.5 indicate a stronger effect values between 0.5 and 0.8 indicate a very strong effect.</p> <p>Cronbach's Alpha of the scale measuring the outcome variable <i>Influence of the Use of Program Theory in the Final Report</i> indicates the scale's reliability is substantial (<math>\alpha = 0.81</math>).</p>						



CORE MODELS	3A, 3B,	INFLUENCE OF INVOLVEMENT IN PROGRAM THEORY ELABORATION
	4A & 4B	IMPORTANCE/LIKELIHOOD OF USE

The fit statistics of the final two pairs of Core Model versions indicate they are not consistent with the data from which they have been estimated. Nevertheless, they have been investigated as part of this study. Therefore, the path diagrams of all four versions with fit statistics have been included below. As can be seen in Table 7.7, none of the four Models fit in accordance with the suggested fit statistic criteria thresholds.

#### SUMMARY

This section has described the testing of the eight versions of the Core Model. Four of these were found to meet the criterion for a fitted model (1a, 1b, 2a, 2b) and diagrams representing each of the versions with fit statistics have been included. The indirect, direct and total effect values have been given for each Model. Model versions 1a and 1b account for 0.46 and 0.47 of the variance in the outcome variable, while Model versions 2a and 2b account for 0.27 and 0.28 of the variance in the outcome variable, respectively, as indicated by the squared multiple correlation values for the models. The fit statistics for all eight model versions are presented in Table 7.7. Further discussion of the findings of the structural equation modelling analysis of the model versions is undertaken in Chapter VIII.

**Table 7.7: Goodness of Fit for the Eight Core Models**

Model	$\chi^2$	DF	$p^1$	CMIN/df <sup>2</sup>	TLI <sup>3</sup>	GFI <sup>4</sup>	AGFI <sup>4</sup>	PGFI <sup>5</sup>	RMSEA <sup>6</sup>
<b>1a</b>	30.46	24	0.17	1.27	0.86	0.85	0.71	0.45	0.09
<b>1b</b>	30.67	24	0.16	1.28	0.86	0.85	0.71	0.45	0.09
<b>2a</b>	33.10	24	0.10	1.38	0.79	0.84	0.69	0.45	0.11
<b>2b</b>	33.44	24	0.10	1.39	0.79	0.83	0.69	0.45	0.11
<b>3a</b>	38.12	24	0.03	1.59	0.78	0.80	0.62	0.42	0.13
<b>3b</b>	41.11	24	0.02	1.71	0.73	0.78	0.59	0.42	0.14
<b>4a</b>	98.88	50	0.00	1.98	0.68	0.71	0.55	0.46	0.17
<b>4c</b>	99.92	50	0.00	2.00	0.68	0.71	0.55	0.46	0.17
<p>Notes:</p> <p><sup>1</sup> A probability value for Chi Square test smaller than 0.05 indicates the null hypothesis that the model fits the data should be rejected.</p> <p><sup>2</sup> Holmes-Smith (2001) suggests CMIN/df values should be greater than 1 but less than 2. However, he considers values between 2 and 3 are to be indicative of reasonable fit while values less than 1 indicate over fit.</p> <p><sup>3</sup> TLI values between 0.90 and 1.0 are considered to be a good fit with the data.</p> <p><sup>4</sup> AGFI and GFI values close to 1 are considered to be indicative of a good fit (Hu and Bentler 1995).</p> <p><sup>5</sup> Parsimony-based index values are often lower than the acceptable threshold for other normed indices of fit. For instance, goodness of fit indices with a value of around 0.90 range can be linked with parsimonious fit indices around .50 (Bryne, 2001; Mulaik et al., 1989).</p> <p><sup>6</sup> The more generous upper threshold (0.1) of Browne and Cudeck (1989) has been used with regards to the RMSEA evaluation of model fit.</p>									

## FURTHER DEVELOPMENT OF THE CORE MODEL

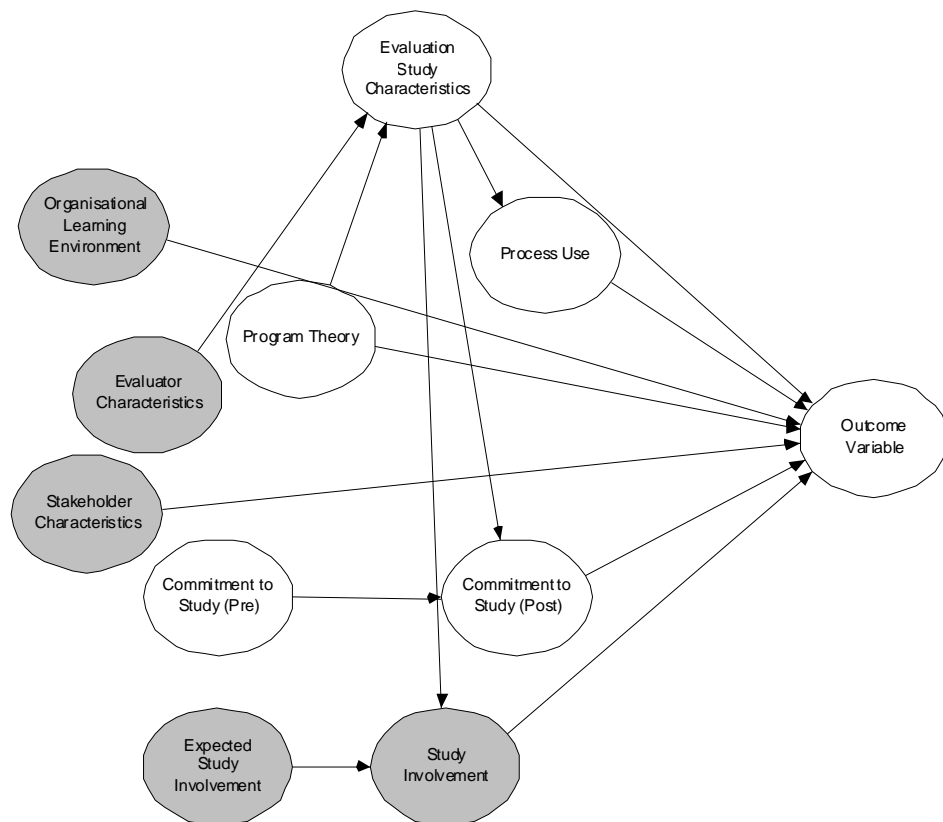
The fully Elaborated Model is presented in Figure 7.11. An analysis of the model indicated that it did not fit the data. Thirty-six co-varies were suggested to improve Model fit. Four of the suggested co-varies with the greatest impact on Model fit were included in a Model re-specification. The re-specified Model, however, also did not fit the data. Therefore, work has been undertaken to further develop the Core Model by adding the additional variables of the fully Elaborated Model one at a time. This section focuses on the exploratory model development phase of the SEM analysis.

The Core Model of the study which was tested included factors prioritised in accordance with a review of relevant literature indicating their influence on an evaluation and the information. In this section, further factors identified in pertinent literature as having an influence on an evaluation's information, are included in the Model. The five additional latent predictor factors included in the Elaborated Structural Model undertaken in this section are: *Organisational Environment Characteristics*, *Evaluator Characteristics*, *Stakeholder Characteristics*, *Expected Study Involvement* and *Study Involvement (Post)*. These variables are shaded in Figure 7.11. The Elaborated Model is tested using the same four outcome variables as used in testing the Core Model. The latent predictor variables included in the Core Model tested above, remain constant.

As the fully Elaborated Model did not fit the data, a strategy to test the Elaborated Model, was undertaken. Three of the latent variables included in the Elaborated Model, but not the Core Model, (*Organisational Environment Characteristics*, *Stakeholder Characteristics*, and *Evaluator Characteristics*) were added one at a time to the Structural Model. The Model was developed and tested in this manner as the addition of two of these variables at the same time resulted in a positive linearity problem which AMOS was not able to overcome to analyse the data. Positive linearity results when knowing the value of one variable enables the prediction of a value with regards to the

second variable. This problem was due to the manner in which the data was managed and resulted primarily from the collection of data from different subject groups (refer to Chapter V).

The inclusion of the latent variables in the Model development has been undertaken in accordance with a temporal sequence logic considered important in the influences on the undertaking of an evaluation study. For instance, the *Organisational Environment Characteristics* is likely to have an influence on the stakeholders of the evaluation, the evaluators chosen to undertake the evaluation and the nature of the evaluation process. Furthermore, the organisational environment determines, to some degree, the opportunity for the use of the evaluation information (Preskill & Torres, 1999a). Finally, *Expected Study Involvement* and *Study Involvement (Post)* have been included jointly as a pair, as they represent pre- and post- measures regarding study involvement. The latent variables added in the development of the Elaborated Model are discussed below.



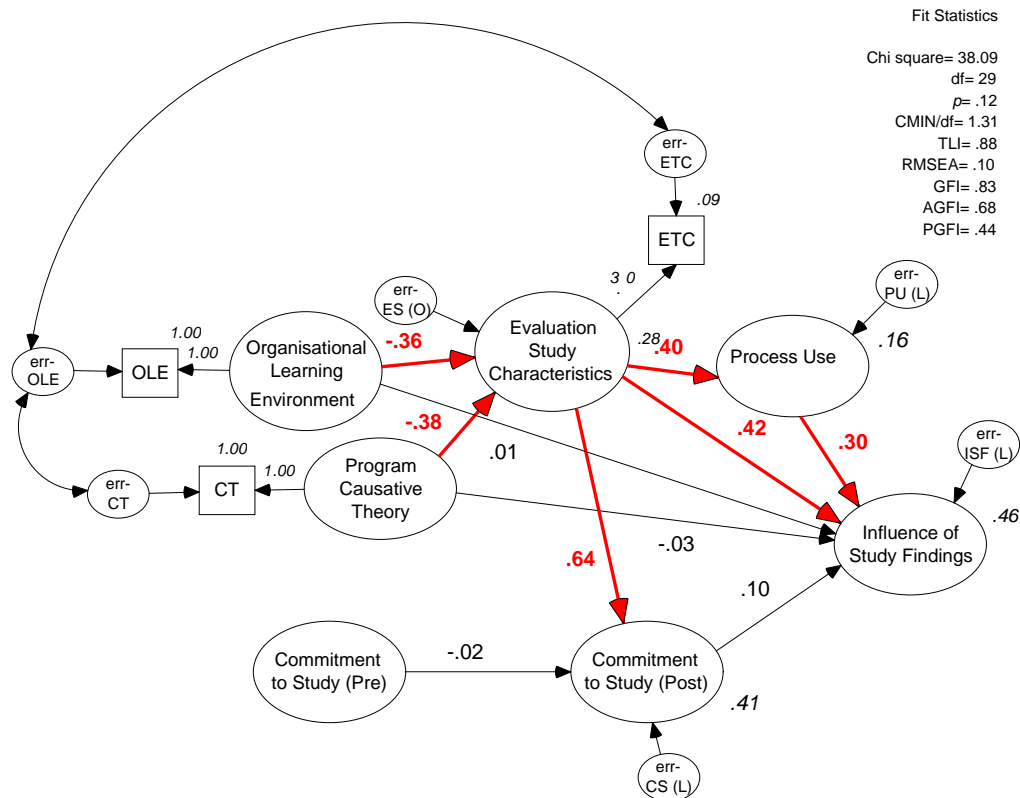
**Figure 7.11: Fully Elaborated Model**

## LATENT VARIABLES ADDED IN THE ELABORATION OF THE CORE MODEL

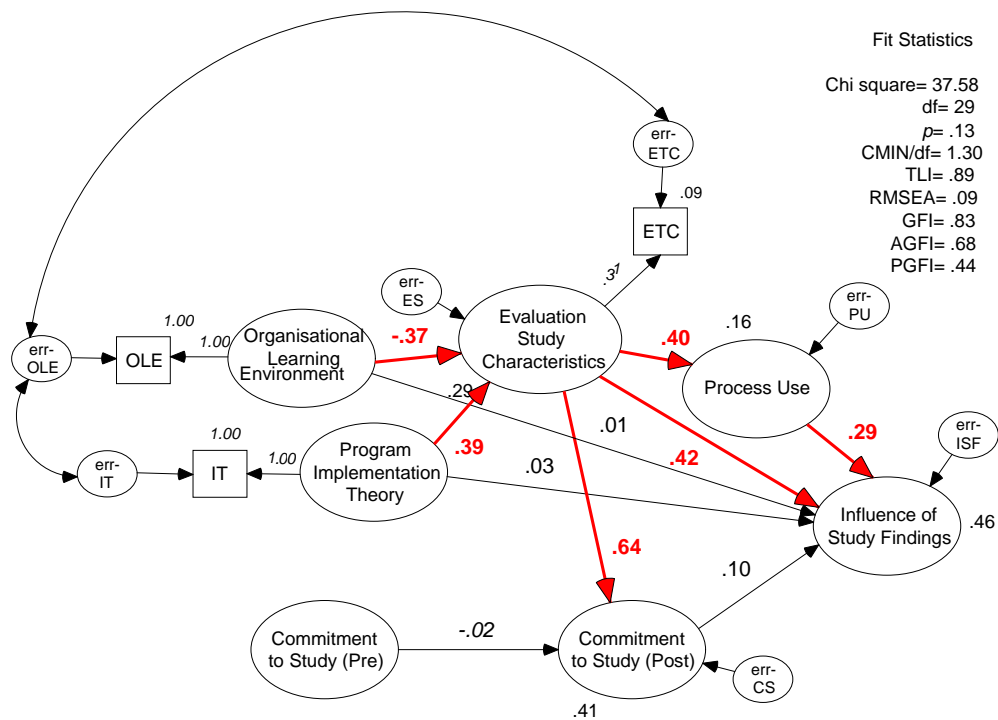
### ORGANISATIONAL LEARNING ENVIRONMENT

The work of Preskill and Torres (1999a, 1999b, 2000a) considers the importance of evaluative inquiry for organisational learning. They argue that for evaluative inquiry to be successful, an organisation must have the infrastructure to support it (1999a). This proposal is pertinent to the present study and has been included in the Elaborated Structural Model.

The Elaborated Model version with the *Organisational Environment Characteristics* latent variable added did not fit the data. The modification indices of the AMOS output, however, indicated co-varying the error terms of *Organisational Environment Characteristics* with *Program Causative Theory*, and the *Organisational Environment Characteristics* error term with Evaluation Team Characteristics (an observed variable indicating the *Evaluation Study Characteristics* latent variable) would improve the Model fit. Co-varying the error terms, in both cases, made sense conceptually. The *Organisational Environment Characteristics* latent variable, with the error terms co-varied for each in the same manner, was introduced into all eight versions of the Core Model. The fit statistics indicate four of the Elaborated Model versions to be a reasonable fit of the data (figures 7.12, 7.13, 7.14, and 7.15). The regression paths, with estimates indicated to be significant at the 0.10 level, have been marked in red on the structural model diagrams. Tables 7.8 and 7.10 present the regression path estimates of core models 1a, 1b, 2a, and 2b. The Direct, Indirect and Total Effect Statistics for Core Model versions 1a and 1b elaborated with the *Organisational Environment Characteristics* variable, are reported in Table 7.9, and those for Elaborated Model versions 2a and 2b have been included in Table 7.11. The fit statistics of all model versions, further elaborated with the *Organisational Environment Characteristics* latent variable, are given in Table 7.12.



**Figure 7.12: Core Model 1a Elaborated with OLE Variable**



**Figure 7.13: Core Model 1b Elaborated with OLE Variable**

**Table 7.8 Regression Weights for Model 1 Elaborated with OLE Variable**

Regression Path Variables		Model 1a Estimate (Probability)	Model 1b Estimate (Probability)
Predictor Variable	Predicted Variable		
Program Causative Theory	Evaluation Study Characteristics	-0.38 (0.03)	
Program Implementation Theory	Evaluation Study Characteristics		0.39 (0.03)
Program Causative Theory	Influence of Study Findings	-0.03 (0.83)	
Program Implementation Theory	Influence of Study Findings		0.03 (0.83)
Organisational Learning Environment	Evaluation Study Characteristics	-0.36 (0.04)	-0.37 (0.03)
Organisational Learning Environment	Influence of Study Findings	0.01 (0.91)	0.01 (0.93)
Evaluation Study Characteristics	Process Use	0.40 (0.02)	0.40 (0.02)
Evaluation Study Characteristics	Commitment to Study (Post)	0.64 ( $< 0.001$ )	0.64 ( $< 0.001$ )
Evaluation Study Characteristics	Influence of Study Findings	0.42 (0.06)	0.42 (0.07)
Process Use	Influence of Study Findings	0.30 (0.03)	0.29 (0.03)
Commitment to Study (Pre)	Commitment to Study (Post)	-0.02 (0.90)	-0.02 (0.90)
Commitment to Study (Post)	Influence of Study Findings	0.10 (0.55)	0.10 (0.56)

**Table 7.9 Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Study Findings with inclusion of Organisational Learning as a Predictor Variable (Model Version 1)**

<b>Theory Predictor :</b>	<b>Causative Theory (Version 1a)</b>			<b>Implementation Theory (Version 1b)</b>		
<i>Squared Multiple Correlation for Structural Equation</i>	<b>R<sup>2</sup>= 0.46</b>			<b>R<sup>2</sup>= 0.46</b>		
	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>
<b>Predictor Variable</b>						
<b>Program Causative Theory</b>	-0.03	-0.23	-0.27			
<b>Program Implementation Theory</b>				0.03	0.24	0.27
<b>Organisational Learning Environment</b>	0.01	-0.22	-0.21	0.01	-0.23	-0.22
<b>Evaluation Study Characteristics</b>	0.42	0.18	0.61	0.43	0.18	0.61
<b>Process Use</b>	0.30	0.00	0.30	0.30	0.00	0.30
<b>Commitment to Study (Pre)</b>	0.00	0.00	0.00	0.00	0.00	0.00
<b>Commitment to Study (Post)</b>	0.10	0.00	0.10	0.10	0.00	0.10
Note: Holmes-Smith (personal communication, July 24, 2007) suggests any effect size value below 0.2 be considered weak, a value between 0.2 and 0.3 be considered small, while values between 0.3 and 0.5 indicate a stronger effect values between 0.5 and 0.8 indicate a very strong effect.						



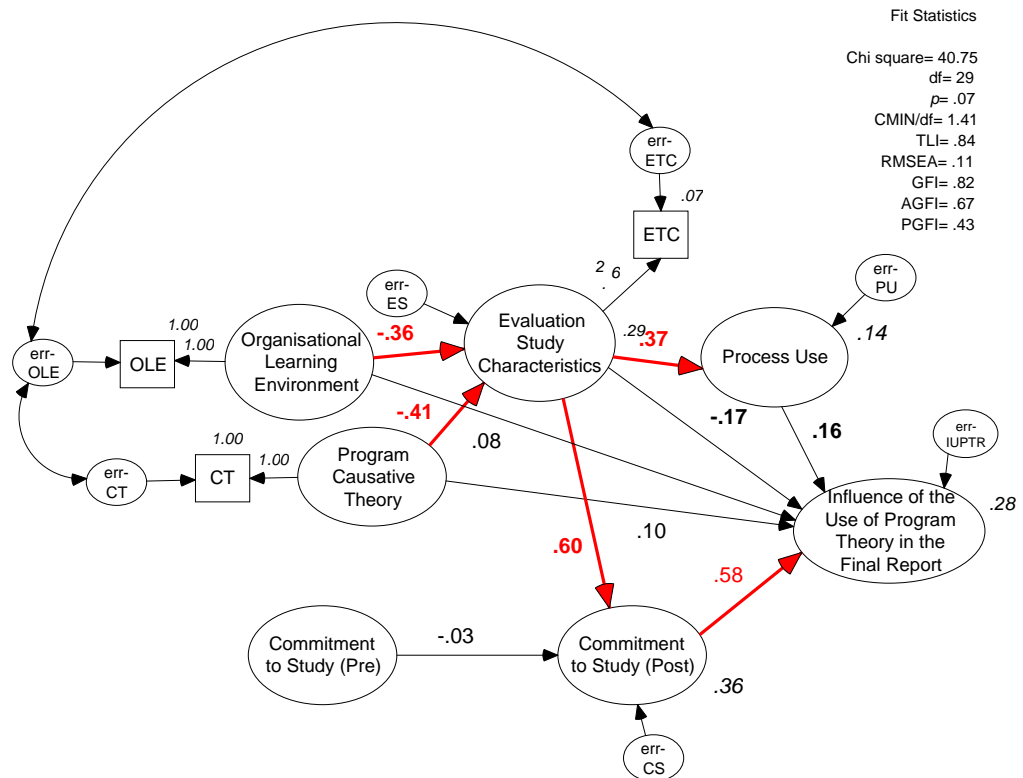


Figure 7.14: Core Model 2a Elaborated with OLE Variable

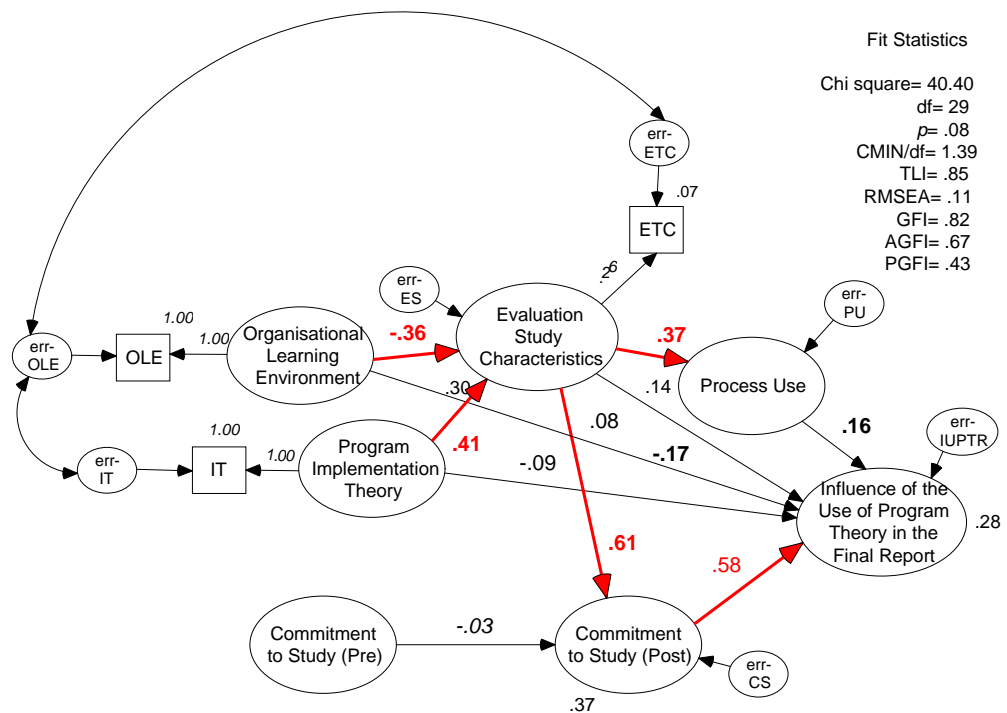


Figure 7.15: Core Model 2b Elaborated with OLE Variable

**Table 7.10 Regression Weights for Model 2 Elaborated with OLE Variable**

Regression Path Variables		Model 2a Estimate (Probability)	Model 2b Estimate (Probability)
Predictor Variable	Predicted Variable		
Program Causative Theory	Evaluation Study Characteristics	-0.41 (0.03)	
Program Implementation Theory	Evaluation Study Characteristics		0.41 (0.02)
Program Causative Theory	Influence of Use of PT in the Final Report	0.10 (0.59)	
Program Implementation Theory	Influence of Use of PT in the Final Report		-0.09 (0.62)
Organisational Learning Environment	Evaluation Study Characteristics	-0.36 (0.04)	-0.36 (0.04)
Organisational Learning Environment	Influence of Use of PT in the Final Report	0.08 (0.58)	0.08 (0.63)
Evaluation Study Characteristics	Process Use	0.37 (0.03)	0.37 (0.03)
Evaluation Study Characteristics	Commitment to Study (Post)	0.60 ( $< 0.001$ )	0.61 ( $< 0.001$ )
Evaluation Study Characteristics	Influence of Use of PT in the Final Report	-0.17 (0.44)	-0.17 (0.44)
Process Use	Influence of Use of PT in the Final Report	0.16 (0.30)	0.16 (0.30)
Commitment to Study (Pre)	Commitment to Study (Post)	-0.03 (0.84)	-0.03 (0.84)
Commitment to Study (Post)	Influence of Use of PT in the Final Report	0.58 ( $< 0.001$ )	0.58 ( $< 0.001$ )

**Table 7.11 Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Use of Program Theory in Final Report with inclusion of Organisational Learning as a Predictor Variable (Core Model Version 2)**

<b>Theory Predictor :</b>	<b>Causative Theory (Version 2a)</b>			<b>Implementation Theory (Version 2b)</b>		
<i>Squared Multiple Correlation for Structural Equation</i>	<b>R<sup>2</sup> = 0.28</b>			<b>R<sup>2</sup> = 0.28</b>		
	Standardised Direct Effects	Standardised Indirect Effects	Standardised Total Effects	Standardised Direct Effects	Standardised Indirect Effects	Standardised Total Effects
<b>Predictor Variable</b>						
<b>Program Causative Theory</b>	0.10	-0.10	0.00			
<b>Program Implementation Theory</b>				-0.09	0.10	0.01
<b>Organisational Learning Environment</b>	0.08	-0.09	0.00	0.08	-0.09	-0.01
<b>Evaluation Study Characteristics</b>	-0.17	0.41	0.24	-0.17	0.42	0.24
<b>Process Use</b>	0.17	0.00	0.17	0.16	0.00	0.16
<b>Commitment to Study (Pre)</b>	0.00	-0.02	-0.02	0.00	-0.02	-0.02
<b>Commitment to Study (Post)</b>	0.58	-0.02	0.58	0.58	0.00	0.58
Note: Holmes-Smith (personal communication, July 24, 2007) suggests any effect size value below 0.2 be considered weak, a value between 0.2 and 0.3 be considered small, while values between 0.3 and 0.5 indicate a stronger effect values between 0.5 and 0.8 indicate a very strong effect.						

**Table 7.12: Goodness of Fit Measures for Core Structural Model Versions Elaborated with Organisational Learning Environment**

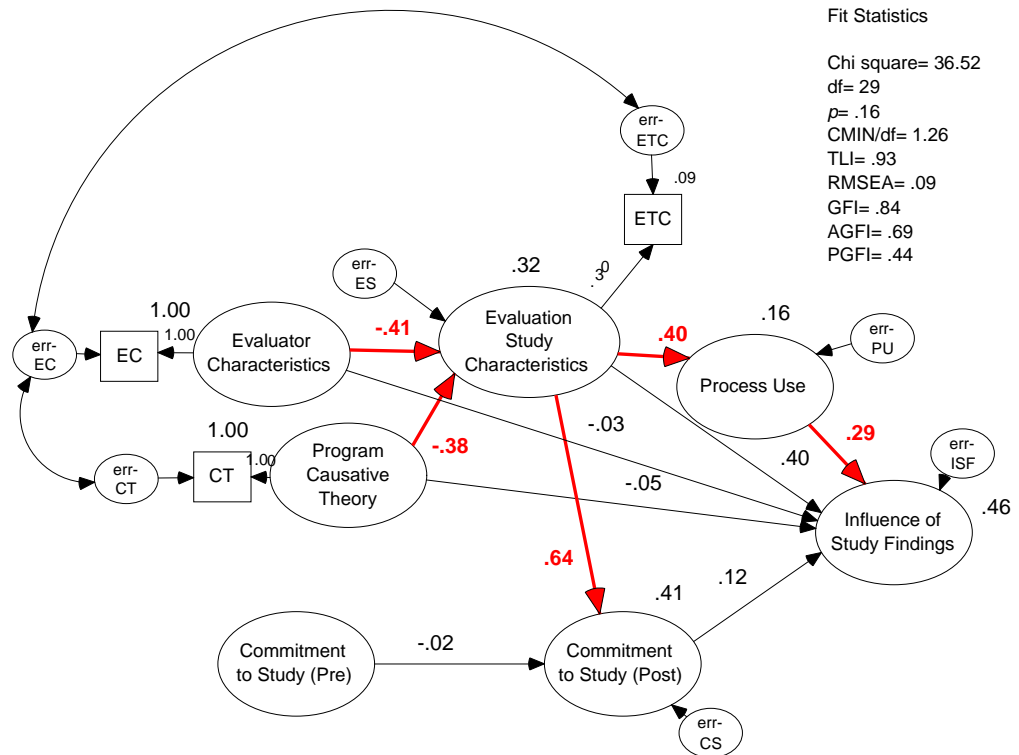
Version	$\chi^2$	df	$p^1$	CMIN/df <sup>2</sup>	TLI <sup>3</sup>	GFI <sup>4</sup>	AGFI <sup>4</sup>	PGFI <sup>5</sup>	RMSEA <sup>6</sup>
<b>1a</b>	38.09	29	0.12	1.31	0.88	0.83	0.68	0.44	0.10
<b>1b</b>	37.58	29	0.13	1.30	0.89	0.83	0.68	0.44	0.09
<b>2a</b>	40.75	29	0.07	1.41	0.84	0.82	0.67	0.43	0.11
<b>2b</b>	40.40	29	08	1.39	0.85	0.82	0.67	0.43	0.11
<p>Notes: <sup>1</sup> A probability value for Chi Square test smaller than 0.05 indicates the null hypothesis that the model fits the data should be rejected.</p> <p><sup>2</sup> Holmes-Smith (2001) suggests CMIN/df values should be greater than 1 but less than 2. However, he considers values between 2 and 3 are to be indicative of reasonable fit while values less than 1 indicate over fit.</p> <p><sup>3</sup> TLI values between 0.90 and 1.0 are considered to be a good fit with the data.</p> <p><sup>4</sup> AGFI and GFI values close to 1 are considered to be indicative of a good fit (Hu and Bentler 1995).</p> <p><sup>5</sup> Parsimony-based index values are often lower than the acceptable threshold for other normed indices of fit. For instance, goodness of fit indices with a value of around .90 range can be linked with parsimonious fit indices around .50 (Bryne, 2001; Mulaik et al., 1989).</p> <p><sup>6</sup> The more generous upper threshold (0.1) of Browne and Cudeck (1989) has been used with regards to the RMSEA evaluation of model fit.</p>									

## EVALUATOR CHARACTERISTICS

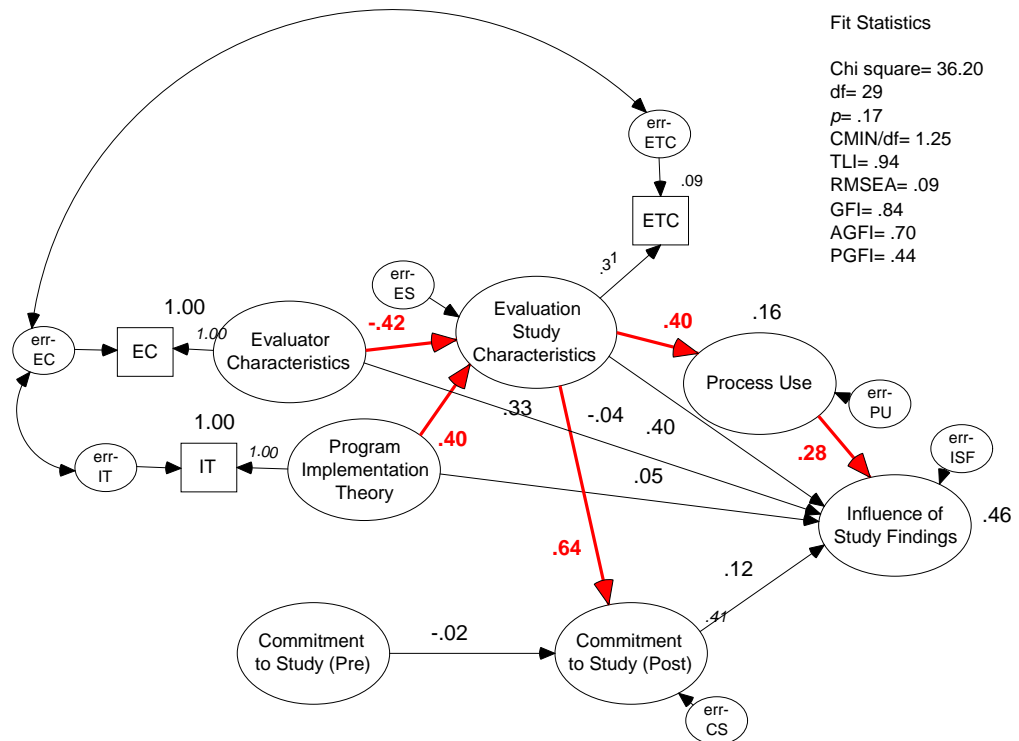
The influence of the characteristics of the evaluators or evaluation team on the evaluation process and outcomes has been written about by many authors (e.g. Braskamp et al., 1982; Conley-Tyler, 2005; Cummings et al., 1988; Greene, 1988; House and Howe, 1998; Lake, 2005; Mathison, 1994; Owen, 2006; Scriven, 1991). Therefore, the influence of *Evaluator Characteristics* as a latent predictor variable in the elaborated version of the Model is of interest.

The *Evaluator Characteristics* latent variable was added to all eight versions of the Core Model. As with the model elaborations including *Organisational Environment Characteristics*, modification indices indicated co-varying the error terms of *Evaluator Characteristics* and *Program Causative Theory*, as well as the *Organisational Environment Characteristics* error term and Evaluation Team Characteristics (an observed variable indicating the *Evaluation Study Characteristics* latent variable) would improve the Model fit. The conceptual considerations of co-varying the error terms are the same as those identified above in the section detailing the Model elaborations including the *Organisational Environment Characteristics* Model. Co-varying the error terms as suggested by the AMOS output improved Model fit.

The Model versions 1a, 1b, 2a, and 2b elaborated with the *Evaluator Characteristics* latent variable fit the data reasonably well. Model versions 1a and 1b elaborated with the *Evaluator Characteristics* latent variable have been included in Figures 7.16 and Figure 7.17. The elaborated versions of Models 2a and 2b have been included in Figure 7.18 and Figure 7.19. Table 7.13 and Table 7.15 present the regression path estimates for the versions of Model 1 and 2 elaborated with the *Evaluator Characteristics* variable. The direct, indirect and total effect statistics for Model versions 1 and 2 elaborated with the *Evaluator Characteristics* variable have been included in Table 7.14 and Table 7.16, respectively. Table 7.17 includes the fit statistics for the elaborated versions 1a, 1b, 2a, and 2b of the Core Model.



**Figure 7.16: Model Version 1a Elaborated with Evaluator Characteristics**



**Figure 7.17: Model Version 1b Elaborated with Evaluator Characteristics**

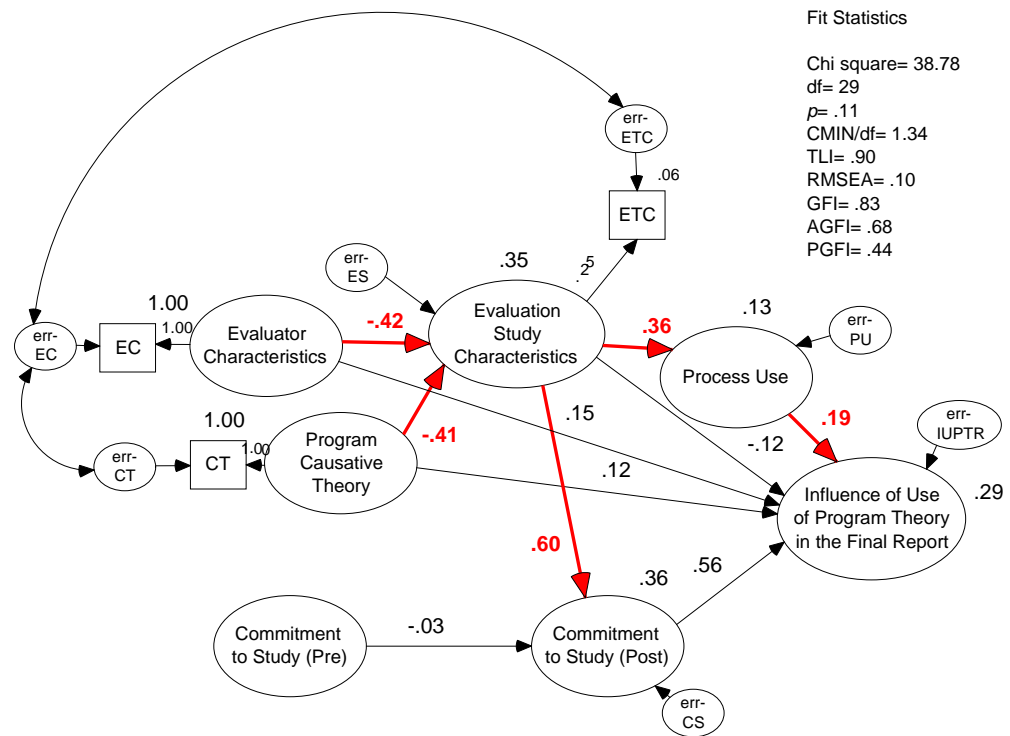
**Table 7:13 Regression Weights for Model Version 1 Elaborated with Evaluator Characteristics**

Regression Path Variables		Model 1a Estimate (Probability)	Model 1b Estimate (Probability)
Predictor Variable	Predicted Variable		
Program Causative Theory	Evaluation Study Characteristics	-0.38 (0.03)	
Program Implementation Theory	Evaluation Study Characteristics		0.40 (0.03)
Program Causative Theory	Influence of Study Findings	-0.05 (0.79)	
Program Implementation Theory	Influence of Study Findings		0.05 (0.78)
Evaluator Characteristics	Evaluation Study Characteristics	-0.41 (0.02)	-0.42 (0.02)
Evaluator Characteristics	Influence of Study Findings	-0.03 (0.84)	-0.04 (0.83)
Evaluation Study Characteristics	Process Use	0.29 (0.02)	0.28 (0.02)
Evaluation Study Characteristics	Commitment to Study (Post)	0.64 ( $< 0.001$ )	0.64 ( $< 0.001$ )
Evaluation Study Characteristics	Influence of Study Findings	0.40 (0.11)	0.40 (0.11)
Process Use	Influence of Study Findings	0.29 (0.04)	0.29 (0.04)
Commitment to Study (Pre)	Commitment to Study (Post)	-0.02 (0.90)	-0.02 (0.90)
Commitment to Study (Post)	Influence of Study Findings	0.12 (0.51)	0.12 (0.51)

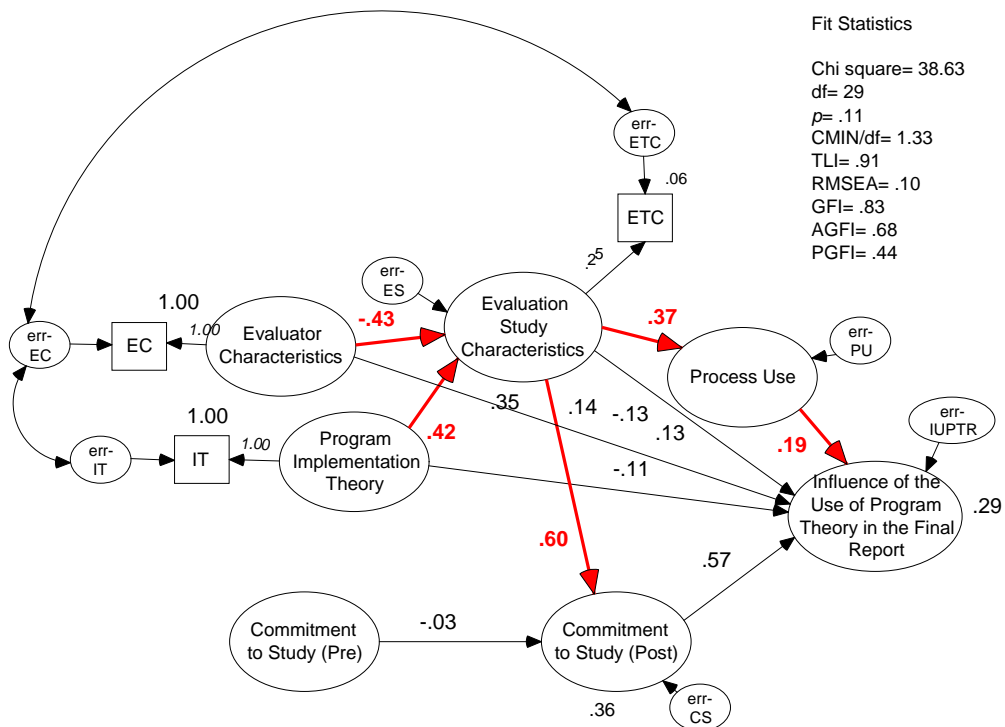
**Table 7.14 Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of Study Findings with inclusion of Evaluator Characteristics as a Predictor Variable (Model Version 1)**

<b>Theory Predictor :</b>	<b>Causative Theory (Version 1a)</b>			<b>Implementation Theory (Version 1b)</b>		
<i>Squared Multiple Correlation for Structural Equation</i>	<b>R<sup>2</sup> = 0.46</b>			<b>R<sup>2</sup> = 0.46</b>		
	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>
<b>Predictor Variable</b>						
<b>Program Causative Theory</b>	-0.05	-0.22	-0.27			
<b>Program Implementation Theory</b>				0.05	0.23	0.30
<b>Evaluator Characteristics</b>	-0.03	-0.24	-0.28	-0.4	-0.24	-0.28
<b>Evaluation Study Characteristics</b>	0.40	0.19	0.58	0.40	0.19	0.58
<b>Process Use</b>	0.30	0.00	0.29	0.29	0.00	0.29
<b>Commitment to Study (Pre)</b>	0.00	0.00	0.00	0.00	0.00	0.00
<b>Commitment to Study (Post)</b>	0.12	0.00	0.11	0.12	0.00	0.12
Note: Holmes-Smith (personal communication, July 24, 2007) suggests any effect size value below 0.2 be considered weak, a value between 0.2 and 0.3 be considered small, while values between 0.3 and 0.5 indicate a stronger effect values between 0.5 and 0.8 indicate a very strong effect.						





**Figure 7.18: Model Version 2a Elaborated with Evaluator Characteristics**



**Figure 7.19: Model Version 2b Elaborated with Evaluator Characteristics**

**Table 7:15 Regression Weights for Model Version 2 Elaborated with Evaluator Characteristics**

Regression Path Variables		Model 2a Estimate (Probability)	Model 2b Estimate (Probability)
Predictor Variable	Predicted Variable		
Program Causative Theory	Evaluation Study Characteristics	-0.41 (0.03)	
Program Implementation Theory	Evaluation Study Characteristics		0.42 (0.02)
Program Causative Theory	Influence of Use of PT in the Final Report	0.12 (0.55)	
Program Implementation Theory	Influence of Use of PT in the Final Report		-0.11 (0.57)
Evaluator Characteristics	Evaluation Study Characteristics	-0.42 (0.02)	-0.43 (0.02)
Evaluator Characteristics	Influence of Use of PT in the Final Report	0.15 (0.45)	0.14 (0.50)
Evaluation Study Characteristics	Process Use	0.36 (0.03)	0.37 (0.03)
Evaluation Study Characteristics	Commitment to Study (Post)	0.60 ( $< 0.001$ )	0.60 ( $< 0.001$ )
Evaluation Study Characteristics	Influence of Use of PT in the Final Report	-0.12 (0.60)	0.13 (0.59)
Process Use	Influence of Use of PT in the Final Report	0.19 (0.22)	0.19 (0.23)
Commitment to Study (Pre)	Commitment to Study (Post)	-0.03 (0.83)	-0.03 (0.83)
Commitment to Study (Post)	Influence of Use of PT in the Final Report	0.57 (0.01)	0.56 (0.01)

**Table 7.16 Direct, Indirect and Total Effects of Latent Predictor Factors on Influence of the Use of Program Theory in the Final Report with Evaluator Characteristics included as a Predictor Variable (Model Version 2).**

<b>Theory Predictor :</b>	<b>Causative Theory (Version 2a)</b>			<b>Implementation Theory (Version 2b)</b>		
<i>Squared Multiple Correlation for Structural Equation</i>	<b>R<sup>2</sup> = 0.29</b>			<b>R<sup>2</sup> = 0.29</b>		
	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>
<b>Predictor Variable</b>						
<b>Program Causative Theory</b>	0.12	-0.12	0.00			
<b>Program Implementation Theory</b>				-0.08	0.09	0.01
<b>Evaluator Characteristics</b>	0.15	-0.12	0.03	0.14	-0.12	0.02
<b>Evaluation Study Characteristics</b>	-0.12	0.41	0.29	-0.13	0.41	0.28
<b>Process Use</b>	0.19	0.00	0.19	0.19	0.00	0.19
<b>Commitment to Study (Pre)</b>	0.00	-0.02	-0.02	0.00	-0.02	-0.02
<b>Commitment to Study (Post)</b>	0.56	0.00	0.56	0.59	0.00	0.57
Note: Holmes-Smith (personal communication, July 24, 2007) suggests any effect size value below 0.2 be considered weak, a value between 0.2 and 0.3 be considered small, while values between 0.3 and 0.5 indicate a stronger effect values between 0.5 and 0.8 indicate a very strong effect.						

**Table 7.17: Goodness of Fit Measures for Core Structural Model Versions with Evaluator Characteristics Latent Variable Added**

Model	$\chi^2$	df	$p^1$	CMIN/df <sup>2</sup>	TLI <sup>3</sup>	GFI <sup>4</sup>	AGFI <sup>4</sup>	PGFI <sup>5</sup>	RMSEA <sup>6</sup>
1a	36.52	29	0.16	1.26	0.93	0.84	0.69	0.44	0.09
1b	36.20	29	0.17	1.25	0.94	0.84	0.70	0.44	0.09
2a	38.78	29	0.11	1.34	0.90	0.83	0.68	0.44	0.11
2b	38.63	29	0.11	1.33	0.91	0.83	0.68	0.44	0.11
<p>Notes: <sup>1</sup> A probability value for Chi Square test smaller than 0.05 indicates the null hypothesis that the model fits the data should be rejected.</p> <p><sup>2</sup> Holmes-Smith (2001) suggests CMIN/df values should be greater than 1 but less than 2. However, he considers values between 2 and 3 are to be indicative of reasonable fit while values less than 1 indicate over fit.</p> <p><sup>3</sup> TLI values between 0.90 and 1.0 are considered to be a good fit with the data.</p> <p><sup>4</sup> AGFI and GFI values close to 1 are considered to be indicative of a good fit (Hu and Bentler 1995).</p> <p><sup>5</sup> Parsimony-based index values are often lower than the acceptable threshold for other normed indices of fit. For instance, goodness of fit indices with a value of around .90 range can be linked with parsimonious fit indices around .50 (Bryne, 2001; Mulaik et al., 1989).</p> <p><sup>6</sup> The more generous upper threshold (0.1) of Browne and Cudeck (1989) has been used with regards to the RMSEA evaluation of model fit.</p>									

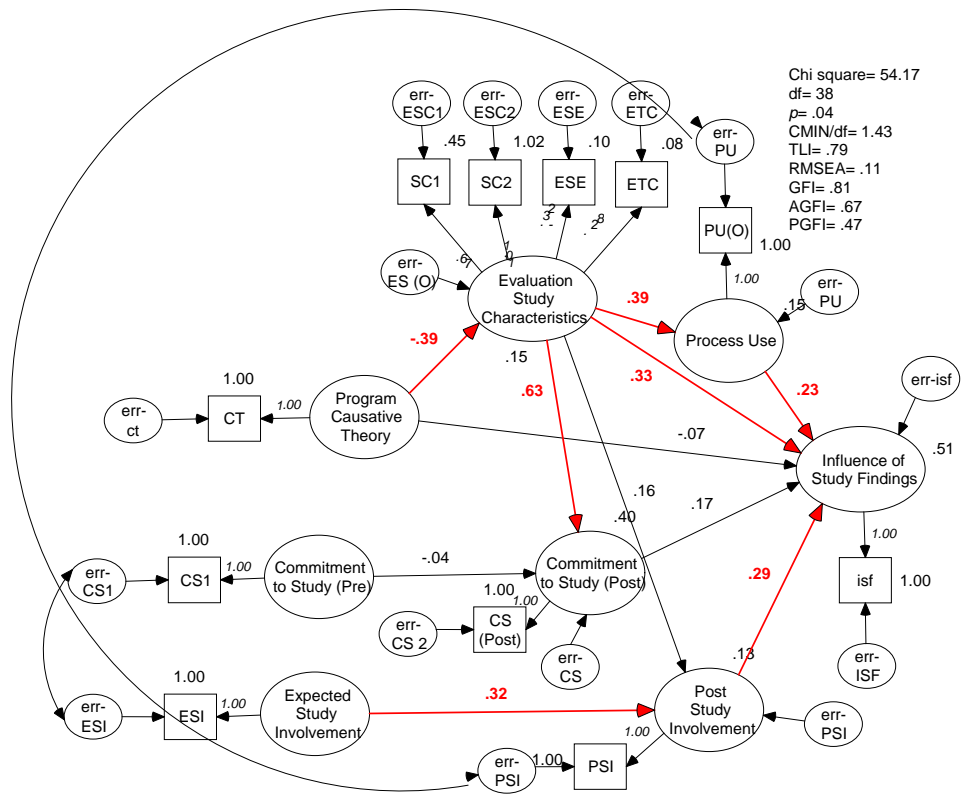
## STAKEHOLDER CHARACTERISTICS

Stakeholder Characteristics have been found by many authors to have an influence on the use of evaluation information (Cousins and Walker, 2000; Cummings, 1997; Hudson-Mabbs, 1993; Leviton and Hughes, 1981; Vlahov, 1990) and so this latent predictor variable is included in the elaboration of the Core Model. Testing of the Core Models of the study elaborated with the addition of the *Stakeholder Characteristics* latent variable, found none fit the data. Modification to the models, as suggested by modification indices, only lead to the necessity for further modifications to improve model fit. In many cases the variables or error terms required to co-vary did not have a logical link, therefore it did not make sense to co-vary them. As the models did not fit, no diagrams or tables with fit statistics have been included.

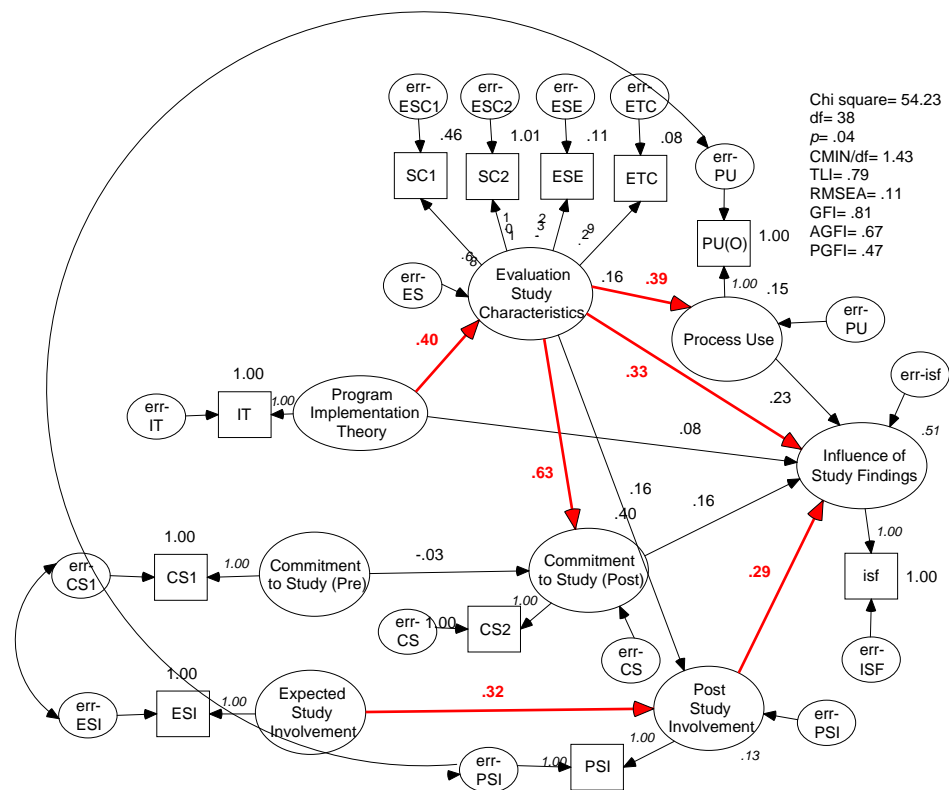
## STUDY INVOLVEMENT

It has been suggested that the involvement of stakeholders in the evaluation prepares them for accepting the evaluation information by making them aware of the program context and the evaluation processes and by increasing their understanding of the evaluation findings, ownership of the evaluation information, sense of personal responsibility for advocating the findings of the study and likelihood of accepting the information as valid and credible (e.g. Cummings, 1997; Preskill and Torres, 1999). Cummings (1997) included items in his instruments assessing both expected study involvement and actual study involvement. *Expected Study Involvement* and *Study Involvement (Post)* have been included as two latent variables in the elaboration of the Structural Equation Model of this study. *Expected Study Involvement* focuses on the expectations of stakeholders regarding their involvement in the evaluation study process, while the *Study Involvement (Post)* factor focuses on the actual involvement of stakeholders in the evaluation process.

Initial attempts to include both latent variables independently in elaborations of the Core Model failed. A modified version including both the *Expected Study Involvement* and *Study Involvement (Post)* variables was then tested. This elaborated version fits the data well, with one co-vary. The error terms of *Study Involvement (Post)* and *Process Use* have been co-varied to facilitate model fit. Conceptually the link makes sense, as both observed variables focus on study involvement. The Model versions 1a and 1b including the *Study Involvement* latent variables (*Expected* and *Actual*) have been included in Figures 7.20 and 7.21 as these two Models have *p* values slightly below the threshold for model fit. Table 7.18 presents the regression path estimates of Structural Models 1a and 1b elaborated with the *Evaluator Characteristics* variable, while the effect statistics for the *Influence of Study Findings* model elaborations have been included in Table 7.19. The fit indices for the elaborations of the Model versions including the *Involvement* latent variables have been included in Table 7.20.



**Figure 7.20: Core Model 1a Elaborated with Involvement Characteristics**



**Figure 7.21: Core Model 1b Elaborated with Involvement Characteristics**

**Table 7:18 Regression Weights for Model Version 1 Elaborated with Expected Study Involvement and Study Involvement (Post)**

Regression Path Variables		Model 1a Estimate (Probability)	Model 1b Estimate (Probability)
Predictor Variable	Predicted Variable		
Program Causative Theory	Evaluation Study Characteristics	-0.39 (0.03)	
Program Implementation Theory	Evaluation Study Characteristics		0.40 (0.03)
Program Causative Theory	Influence of Study Findings	-0.07 (0.60)	
Program Implementation Theory	Influence of Study Findings		0.08 (0.55)
Expected Study Involvement	Study Involvement (Post)	0.32 (0.02)	0.32 (0.02)
Study Involvement (Post)	Influence of Study Findings	0.29 (0.02)	0.29 (0.02)
Evaluation Study Characteristics	Process Use	0.39 (0.03)	0.39 (0.03)
Evaluation Study Characteristics	Commitment to Study (Post)	0.63 ( $< 0.001$ )	0.63 ( $< 0.001$ )
Evaluation Study Characteristics	Study Involvement (Post)	0.16 (0.30)	0.16 (0.30)
Evaluation Study Characteristics	Influence of Study Findings	0.33 (0.10)	0.33 (0.09)
Process Use	Influence of Study Findings	0.23 (0.10)	0.23 (0.10)
Commitment to Study (Pre)	Commitment to Study (Post)	-0.04 (0.73)	-0.03 (0.75)
Commitment to Study (Post)	Influence of Study Findings	0.17 (0.30)	0.16 (0.32)

**Table 7.19 Direct, Indirect and Total Effects of Latent Predictor Factors on the Influence of Study Findings with inclusion of Involvement Predictor Variables (Expected and Actual) (Model Version 1).**

<b>Theory Predictor :</b>	<b>Causative Theory (Version 1a)</b>			<b>Implementation Theory (Version 1b)</b>		
<i>Squared Multiple Correlation for Structural Equation</i>	<b>R<sup>2</sup> = 0.50</b>			<b>R<sup>2</sup> = 0.51</b>		
	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>	<b>Standardised Direct Effects</b>	<b>Standardised Indirect Effects</b>	<b>Standardised Total Effects</b>
<b>Predictor Variable</b>						
<b>Program Causative Theory</b>	0.07	0.23	0.30			
<b>Program Implementation Theory</b>				0.08	0.23	0.31
<b>Expected Study Involvement</b>	0.00	0.09	0.09	0.00	0.09	0.09
<b>Actual Study Involvement</b>	0.29	0.00	0.30	0.29	0.00	0.29
<b>Evaluation Study Characteristics</b>	0.33	0.24	0.57	0.33	0.24	0.57
<b>Process Use</b>	0.23	0.00	0.23	0.23	0.00	0.23
<b>Commitment to Study (Post)</b>	0.17	0.00	0.17	0.16	0.00	0.16
<b>Commitment to Study (Pre)</b>	0.00	0.01	0.01	0.00	0.01	0.01
Note: Holmes-Smith (personal communication, July 24, 2007) suggests any effect size value below 0.2 be considered weak, a value between 0.2 and 0.3 be considered small, while values between 0.3 and 0.5 indicate a stronger effect and values between 0.5 and 0.8 indicate a very strong effect.						



**Table 7.20: Goodness of Fit Measures for Elaborated Model Version 1a and 1b with Involvement Variables (Expected and Actual) Added**

Outcome Variable	$\chi^2$	df	$p^1$	CMIN/df <sup>2</sup>	TLI <sup>3</sup>	GFI <sup>4</sup>	AGFI <sup>4</sup>	PGFI <sup>5</sup>	RMSEA <sup>6</sup>
<b>Influence of Study Findings with Causative Theory</b>	54.17	38	0.04	1.43	0.79	0.81	0.67	0.47	0.11
<b>Influence of Study Findings with Implementation Theory</b>	54.23	38	0.04	1.43	0.79	0.81	0.67	0.47	0.11
<p>Notes: <sup>1</sup> A probability value for Chi Square test smaller than 0.05 indicates the null hypothesis that the model fits the data should be rejected.</p> <p><sup>2</sup> Holmes-Smith (2001) suggests CMIN/df values should be greater than 1 but less than 2. However, he considers values between 2 and 3 are to be indicative of reasonable fit while values less than 1 indicate over fit.</p> <p><sup>3</sup> TLI values between 0.90 and 1.0 are considered to be a good fit with the data.</p> <p><sup>4</sup> AGFI and GFI values close to 1 are considered to be indicative of a good fit (Hu and Bentler 1995).</p> <p><sup>5</sup> Parsimony-based index values are often lower than the acceptable threshold for other normed indices of fit. For instance, goodness of fit indices with a value of around 0.90 range can be linked with parsimonious fit indices around 0.50 (Bryne, 2001; Mulaik et al., 1989).</p> <p><sup>6</sup> The more generous upper threshold (0.1) of Browne and Cudeck (1989) has been used with regards to the RMSEA evaluation of model fit.</p>									

## CONCLUSION

The sections of this chapter detail the steps undertaken in the structural equation modelling process of this study. The measurement characteristics of the Model are analysed first, and acceptable models were found for each latent variable. Structural equation modelling allows the researcher to examine the fit of observed variables on a latent variable predicted by multiple observed variables. The latent variables of the Measurement Model are *Organisational Environment Characteristics*, *Evaluator Characteristics*, *Stakeholder Characteristics*, *Evaluation Study Characteristics* and *Importance/ Likelihood of Use*. The scales of all other latent variables that are predicted by a single scale are presented in Chapter V. The Structural Model of the study is analysed in the second portion of this chapter.

A total of eight versions of the Core Model were tested. A pair of versions of the Model had been developed for each of the four outcome variables of the study which are *Influence of Study Findings*, *Influence of Use of Program Theory in the Final Report*, *Influence of Involvement in the Elaboration of Program Theory*, and *Importance/Likelihood of Use*. Each pair of Model versions had been developed incorporating the *Program Causative Theory* latent variable or the *Program Implementation Theory* latent variable. Only four of the eight versions were found to fit the data according to the fit statistic criteria. These involved the pairs including the *Influence of Study Findings* and the *Influence of the Use of Program Theory in the Final Report*. Details of the analysis of the fit of each version of the Core Model to the data from which they were estimated, are then presented. The versions which include *Influence of Study Findings* as an outcome variable fit the data well. The Core Model versions which include the *Influence of Use of Program Theory in the Final Report* outcome variable are also close to a reasonable fit with the data. The remaining four versions of the Core Model did not fit the data well enough.

Model version 1a with the *Program Causative Theory* and Model version 1b with the *Program Implementation Theory* each explained 46% of the variance in the outcome variable *Influence of Study Findings*. This was considerably

more than for either of the versions with the *Influence of the Use of Program Theory in the Final Report* as the outcome variable (27% and 28%, respectively).

The results of the exploratory analysis of further elaborations of the Core Model are presented in the final sections of the chapter,. The latent variables included in further elaboration of the model include *Organisational Characteristics*, *Evaluator Characteristics*, *Stakeholder Characteristics*, *Expected Study Involvement* and *Study Involvement (Post)*. Although the elaborations including *Stakeholder Characteristics* did not prove to be a good fit, some versions including the other latent variables proved to be a reasonable fit with the data from which they have been estimated. Furthermore, the elaboration of the Core Model versions 1a and 1b with the *Expected Study Involvement* and *Study Involvement (Post)* variables (Figure 7.20 and Figure 7.21) increased the amount of variance accounted for in the outcome variable from 0.45 and 0.46, respectively, to 0.51 for both. The interpretation of the findings of this chapter is presented in the following chapter.

## CHAPTER VIII

### THE EFFECT OF PROGRAM THEORY ON EVALUATION INFORMATION USE

The main purpose of this study was to investigate the singular and interactive effects of identified primary factors on the use of an evaluation's information. The effect of program theory in undertaking an evaluation study on the use of evaluation information, is the major factor of interest in this study. A Core Model (see Chapter III) was developed which includes both this factor and other primary latent predictor factors, identified through previous research to have a substantial impact on the use of an evaluation's information. The logic of this conceptual model underpinned the design, methods, instrument development, data analysis, and structural model development for the study. Further development of the Core Model was undertaken by developing an Elaborated Core Model in which additional latent predictor variables were added and tested, with a view to assessing their impact.

The five variables included in the Core Model are: the type and degree of *program theory* used the evaluation study; the *process use* of the evaluation; the characteristics of the *evaluation study*; and *commitment* to the evaluation study, measured both before, and following, the conduct of the evaluation study. Four types of outcomes were identified for the Core Model, specifically: the influence of *study findings*; the influence of stakeholder involvement in the *program theory elaboration* process; the influence of *use of program theory in the final report*; and, the *use of evaluation information*. A pair of models was developed for each outcome variable, one including *Program Causative Theory* as the program theory predictor variable, and the other including *Program Implementation Theory* as the program theory predictor variable. For instance, the outcome variable of Model 1 is *The Influence of Study Findings*, while the outcome variable of Model 2 is the *Influence of the Use of Program Theory in the Final Report*. The *Program Theory* latent variable of Model 1a and Model 2a is *Program Causative*

*Theory*, while the *Program Theory* latent variable of Model 1b and Model 2b is *Program Implementation Theory*.

This chapter presents a discussion of the findings of the study based upon both the qualitative and the quantitative data gathered in the study with a view to determining the factors which have an effect on the use of the evaluation information. In the first section, the results of testing the four pairs of the Core Model, one pair for each outcome variable, are considered. The findings of the statistical analysis of each model are discussed in turn with consideration also being given to the relevant qualitative information gathered in the study.

In the second part of this chapter the findings for the *Elaborated Core Model* developed in the study are considered. The variables added to the Core Model in its elaboration are: the *learning environment* of the organisation, the characteristics of the *evaluators*, the characteristics of the *stakeholders*, the stakeholders' expectations of *involvement* in the evaluation study, and the *actual involvement* of the stakeholders. The effect on the Model brought about by the inclusion of each variable is considered in turn, followed by a summary of the influence of the variables included in the Elaborated Core Model. The final section of the chapter summarises the findings of the study.

## THE CORE MODEL

An interpretation and discussion of the findings of the statistical analysis for each of the Core Models tested in the study by the four outcome variables is presented below. The Core Model pairs with outcome variables *Influence of Study Findings* and *Influence of the Use of Program Theory* were the only models found to have a reasonable, or close, fit to the data. Relevant qualitative information gathered in the study has been included as appropriate to contextualise the findings and provide further insights. A summary bringing together the findings of the models follows.

## OUTCOME: INFLUENCE OF STUDY FINDINGS

As reported in Chapter VII, the two model pairs with *Influence of Study Findings* as the outcome measure yield the strongest fit statistics indicating the models have an acceptable fit with the data. Although the TLI, GFI, AGFI and PGFI values for the models fell a little below the suggested acceptable threshold the probability ( $p$ ) of calculating the chi square value of both models was above 0.05, the minimum discrepancy values are within the acceptable range and the RMSEA values below the recommended threshold. The  $p$  (chi square) value is commonly considered to be the critical statistic, and should be the first considered in an assessment of model fit. A review of the regression path estimates, and of the standardised direct, indirect and total effects of the latent predictor variables on the outcome, give some insight regarding the links between the variables. The effect values of all the predictor variables, with the exception of the program theory variable, are essentially the same in both models of the pair. Finally, the squared multiple correlation values of the two models indicate that the predictor variables of the models account for 0.45 (Model 1a) and 0.46 (Model 1b) of the variance in the outcome variable, *Influence of Study Findings*. These squared multiple correlation values are considered to be quite high in terms of representing the predictive strength of the models (Holmes-Smith, 2007). Furthermore, the Cronbach's Alpha of the scale measuring the outcome variable *Influence of Study Findings*, also reported in Chapter V, indicates the scale has substantial reliability ( $\alpha = 0.76$ ).

A consideration of the mean score for each evaluation gives some insight into the differences between the impact of the evaluation findings, as measured by the *Influence of Study Findings* outcome variable. The mean scores for the *Influence of Study Findings* scale, the measure for the *Influence of Study Findings* outcome variable, indicates that the YOHFest Program stakeholders (mean score 8.1) are the least influenced by the study findings, while the Pre-Apprenticeship Program stakeholders (mean score 10.8) were most influenced, and the Making Consistent Judgments Program stakeholders (9.9) fell between the two.

*Program Causative Theory, Model 1a*. The statistical tests on the regression estimate values indicate that the use of *Program Causative Theory* as measured in this study has no effect on either the *Evaluation Studies Characteristics* variable or the outcome variable, *Influence of Study Findings*. This finding is at odds with the suggestion of many authors (e.g. Bickman, 1987, 1998; Chen, 1989, 1990a; Pawson and Tilley, 1994; Petrosino 2000; Rogers, 2000) that fuller use of causative theory in undertaking an evaluation study will lead to a greater use of the evaluation information, in that it illuminates both the cause and effect relationships and the nature of the program. In fact, the only two paths of Structural Model 1a indicated as significant by the fit of the regression estimate values, are those linking *Evaluation Study Characteristics* to the *Commitment to the Study (Post)* variables and *Process Use* to the outcome variable *Influence of Study Findings* (refer to Figure 7.7).

A review of the responses of the stakeholders of the evaluation classified as highest on the use of causative theory, regarding the evaluation process, provided some insight. In Chapter VI, the theory classification undertaken for the three evaluation cases rated the YOHFest Program evaluation highest (rating score of 12) on the use of causative theory and the Making Consistent Judgments (MCJ) Program second (rating score of 10). Therefore, a review of the responses of stakeholders of these evaluations to both the qualitative and the quantitative items, provides some insight into the relationship between the *Program Causative Theory* latent variable and the *Evaluation Characteristics* latent variable.

YOHFest Program stakeholders were very happy with the program theory elaboration process they were led through at the beginning of the evaluation study, and stated that they valued the process in terms of the insights they gained into the logic underpinning the program. However, some primary stakeholders of the YOHFest Program were somewhat disillusioned with other aspects of the evaluation process. They thought that the evaluation was going to provide them with both pre- and post- festival data regarding shifts in the participants' thinking about healthy lifestyle issues. In fact, the post-study

collection of empirical data did not occur. Therefore, the empirical data they expected and planned to use in securing funding for the next round of program delivery was not available. Furthermore, many of the stakeholders felt that the distribution of the feedback sheets to the participants post-festival could have been better timed. Finally, the evaluation report was nearly a year overdue before it was available. These issues in particular, led some stakeholders to respond negatively to some of those scale items which measured the observed variables indicating *Evaluation Characteristics*. Furthermore, these same issues were identified by stakeholders as having a limiting and, in some cases, a detrimental impact on the influence of the study findings.

Similarly, MCJ Program stakeholders were very happy with the use of both causative and implementation program theory in the conduct of the evaluation study. However, although staff located within the Curriculum Directorate in the head office of the Department of Education and Training (DET) had a clear avenue through which to put the information to use, those stakeholders located outside the Curriculum Directorate did not. The MCJ Program was centrally developed and driven from the DET head office, but it was delivered at the district level. While feedback from the district offices is considered below in the discussion of the further program development, the central office Curriculum Directorate staff had a clearer vision and greater opportunity to put the information to use.

Both the central and district office staff had good access to the evaluation study report, but some of the stakeholders located elsewhere had to request a copy of the final report to read before they could be interviewed for this study. Thus, many had little time to process the information and are likely to have had little incentive to do so, as they perceived themselves to be too distant from the decision-making processes to have a significant effect on these decisions. Nevertheless, many of them stated that they would be responsible for following through on decisions made at the central office level. Therefore, although some stakeholders of the MCJ Program evaluation responded very positively to the *Influence of Study Findings* scale items, others did not due to their perceived lack of opportunity to put the study findings to use. Finally, the



delivery of the next generation of the MCJ Program had already begun before the release of the final report of the evaluation study and hence the most likely avenue for the evaluation to have an effect was superseded.

*Program Implementation Theory, Model 1b.* The use of program implementation theory in the undertaking of the evaluation study seems to have had a small positive effect (total effect= 0.28) on the *Influence of Study Findings* outcome. The direct effect (0.04) is negligible. Hence, the impact is almost entirely indirect and by way of the *Evaluation Study Characteristics* (indirect effect= 0.24) variable. The standardised regression value of *Program Implementation Theory* on *Evaluation Study Characteristics* is 0.39, a value significant at the 0.05 level (refer to Figure 7.7). This value indicates that a greater use of implementation program theory in the undertaking of an evaluation study is associated with more positive stakeholder perceptions of the characteristics of the evaluation study, such as the environment of the evaluation study process, the competency and practice of the evaluation team, and the quality, timeliness, relevance, nature, and appropriateness of the study findings, as measured in this study. This confirms the claims made by numerous authors (e.g. Chen, 1989; Funnell, 1997; Lipsey and Pollard, 1989; McGaw, 1996; Scheirer, 1987; Scheirer and Rezmovic, 1983; Weiss, 1997) that use of a program's implementation theory focuses the evaluation study on issues pertinent to why the program was, or was not, delivered as intended, enabling valuable insight into program delivery.

A review of the responses of the stakeholders of the evaluation classified in Chapter VI as highest on the use of implementation theory in the undertaking of the evaluation study, provides some further insight. The Pre-Apprenticeship Program was judged to have had the greatest use of Implementation Theory. The responses of the Pre-Apprenticeship Program stakeholders to the qualitative items which focused on the purpose of the evaluation study, indicate they all were very keen to understand what was occurring in the delivery of the program, to gain insight regarding the extent to which the program was meeting industry needs, and to learn what were the outcomes of the Pre-Apprenticeship Program. Changes in the number of hours allocated for

the Pre-Apprenticeship Program based upon a previous evaluation study, had been followed by a meaningful increase in the number of students enrolled in the program. This shift was an important factor in encouraging the stakeholders to undertake the evaluation study. The stakeholders were interested in an objective account of program delivery with a view to gaining an understanding of the program as it was actually being implemented, as opposed to how it was originally intended to be implemented. Although there was some concern expressed regarding the positive nature of the final report of the Pre-Apprenticeship Program evaluation study, the stakeholders generally found the insight gained into the program useful in that it encouraged informed discussion regarding various issues relevant to program delivery which previously had been unclear. The use of the information of the Pre-Apprenticeship Program evaluation, however, was limited, as the program was delivered centrally, but implemented locally. Therefore, TAFE staff responsible for the delivery of the educational components of the program to the young people enrolled in the program, had limited opportunity to put the information to use.

The other evaluation study classified as having included some use of program implementation theory in the design and conduct of the study, is the evaluation of the Making Consistent Judgments Program. The manner in which the MCJ program was implemented was very different to that of previous professional development programs delivered by the Department of Education and Training and hence the effect of the program's implementation on the learning of the school teachers involved, was a primary concern of the evaluation study. Interestingly, with regards to the MCJ Program evaluation, because the implementation of the program was underpinned by social science theory in terms of its adoption of action learning principles, it is difficult to separate the impact of the use of program implementation theory from that of program causative theory. Nevertheless, the MCJ Program stakeholders generally were quite satisfied with the evaluation information regarding program implementation and the feedback from the program participants regarding the effect of the program modules. Furthermore, comments from the stakeholders

regarding their perception of the characteristics of the evaluation were generally quite positive.

*Evaluation Study Characteristics.* Regression estimate values of Model 1a indicate only the path linking *Evaluation Study Characteristics* and *Commitment to the Study (Post)* to be significant ( $p \leq 0.10$ ). This latter effect does not follow through to the outcome variable, *Influence of Study Findings* as the path between *Commitment to the Study (Post)* and the outcome variable is not significant. The link between the *Evaluation Study Characteristics* variable and the outcome variable also is not important as indicated by the regression estimate values. The probability values associated with the regression paths from the *Evaluation Study Characteristics* to the *Process Use* variable, the *Commitment to the Study (Post)* variable and the *Influence of Study Findings* outcome variable, add further insight. With regards to Model 1a none of these pathways are significant. It is suspected that this is due, at least in part, to issues (detailed in the sections above) with the evaluation process specific to the YOHFest Program evaluation; the evaluation study classified highest on the use of program causative theory included in this investigation.

The *Evaluation Study Characteristics* variable plays a more meaningful role in Model 1b. Regression estimate values of the analysis from the Model indicate the *Evaluation Study Characteristics* variable does have links with the *Process Use* variable, *Commitment to the Study (Post)* variable and the outcome variable, *Influence of Study Findings*. Furthermore, the *Evaluation Study Characteristics* variable has a strong effect on the *Influence of the Study Findings* outcome variable as indicated by the indirect (0.42), direct (0.19) and total effect (0.60) values. It appears that the more positively the stakeholders perceive the evaluation study with regards to characteristics such as the environment of the evaluation study process, the competency and practice of the evaluation team, and the quality, timeliness, relevance, nature, and appropriateness of the study findings, the more likely they are to use the study findings.

Generally, the evaluation literature supports these findings. For instance, Cousins and Leithwood (1986) concluded that, based on a variety of research, the variable they identified as 'Evaluation Quality' was the most important factor impacting on the use of an evaluation's information as measured by their "Prevalence of Relationship Index" (p. 349). Similarly, Siegel and Tuckel (1985) found that a negative perception of evaluation methodology was often given as a justification for the dismissal of the evaluation findings, a finding supported by Leviton and Hughes (1981). Generally, it seems that when the characteristics of an evaluation study such as study methodology are perceived positively by stakeholders, the likelihood that the evaluation's information will have influence is increased. This is also supported by the findings of the present study.

*Commitment to the Study (Pre).* The *Commitment to the Study (Pre)* variable has no significant relationship with the outcome variable, as indicated by the regression estimate values. In fact, due to the structure of the Model, *Commitment to the Evaluation (Pre)* variable can affect the outcome variable indirectly through *Commitment to the Program (Post)*. The *Commitment to the Evaluation (Pre)* and *Commitment to the Program (Post)* variables, in the case of both Model 1a and Model 1b, was found not to have a predictive influence on the outcome variable.

*Commitment to the Study (Post).* The regression estimate values of both structural models 1a and 1b indicate that the *Commitment to the Study (Post)* variable is not a predictor of the outcome variable. Therefore, the findings with regards to these two models do not support those of Cummings (1997), who found that commitment to the evaluation defined as involvement, had a greater direct than indirect effect on the *Intention to Use* variable which was the main outcome variable for his study.

A review of the qualitative responses gives some insight into why these results may have occurred. Although the YOHFest Program stakeholders were fairly heavily involved with the evaluation process, some disillusionment with the data collected in the study and the long delay in the writing and completion of

the final report is likely to have led to a reduced use of the evaluation and its findings. With regards to the MCJ Program, although the stakeholders had a reasonable amount of involvement, particularly those located within the Curriculum Directorate in the central office of the Department of Education and Training, delivery of the next round of the MCJ Program had already begun before the report was available. Therefore, the most likely avenue for the evaluation to have influence was superseded. Finally, although the Pre-Apprenticeship Program evaluation appears to have had a reasonable effect on the thinking of the stakeholders, some concerns regarding the manner in which the data were managed caused them to view the findings with caution. Also, alternative programs, addressing the training of apprentices, were already being considered by the Department of Education and Training.

*Process Use.* The regression estimate values and the direct effect values (0.28 and 0.27, respectively) indicate that the *Process Use* variable had a small, yet significant, impact on the outcome variable, *Influence of Study Findings*, in both model 1a and 1b. This finding indicates that the greater the impact of involvement on the thinking, practice and decisions of stakeholders, the greater the effect on the outcome of the evaluation as measured by the *Influence of Study Findings* variable. This finding confirms those of Patton (1997).

## SUMMARY

A review of the relevant regression estimates of the Core Model 1a indicates that including Program Causative Theory in undertaking an evaluation study appears to have little if any, effect with regards to the outcome variable, *Influence of Study Findings*. A consideration of the relevant qualitative data gives some insight as to why this finding may be particular to the cases of this study.

The results with regards to the use of *Program Implementation Theory* are, however, far more positive. The analysis of the structural equation, Core Model 1b indicates that the use of program implementation theory in

undertaking of an evaluation study has a meaningful, if small, positive direct effect on stakeholder's perceptions of the *Evaluation Study Characteristics*, and an indirect positive effect on the outcome variable, *Influence of the Study Findings*.

The *Evaluation Study Characteristics* latent predictor variable has a significant that small positive impact on stakeholder's commitment to the study (post) in both structural model 1a and 1b. In Structural Model 1b, however, regression estimate values indicate that *Evaluation Study Characteristics* also has a meaningful strong effect on the outcome variable and the *Process Use* variable. Furthermore, regression estimates indicate that the *Process Use* variable of Model 1b has a small direct effect on the outcome variable.

*Evaluation Study Characteristics* also has a meaningful small impact on the stakeholder's *Commitment to the Evaluation Study (Post)*. The *Commitment to the Evaluation Study (Pre)* and *Commitment to the Evaluation Study (Post)* variable of both models 1a and 1b, however, do not have an important effect on the outcome variable. Generally, the findings with regards to the Structural Core Model 1b, as detailed earlier in this section, are supported by the relevant works of other authors. The findings with regards to Model 1a tend not to be in agreement with the relevant suggestions of other authors, however. Generally, it is suspected that this is due to difficulties with the evaluation process identified by stakeholders of the YOHFest Program evaluation, as detailed above. Finally, while the works of other authors indicate the *Commitment to the Evaluation Study (Post)* variable has significant effect on the use of the evaluation information, this has not been the case in the analysis of Model 1a and Model 1b. Some possible reasons for the occurrence of this finding have been identified in the relevant discussion above.

## **OUTCOME VARIABLE: INFLUENCE OF THE USE PROGRAM THEORY IN THE FINAL REPORT**

The fit statistics for the models with the outcome variable *Influence of the Use of Program Theory in the Final Report* indicate that Model 2a (refer Figure 7.9) and Model 2b (refer Figure 7.10) are an acceptable fit with the data while the TLI, GFI, AGFI and PGFI values for the models fall a little below the suggested acceptable threshold; the  $p$  value of both models is above 0.05, the minimum discrepancy values are within the acceptable range and the RMSEA values are borderline satisfactory. As mentioned in the above section, the  $p$  value of the chi square estimate is considered here to be a primary critical statistic considered in assessment of model fit. A discussion of the standardised direct, indirect and total effects of the latent predictor variables and the regression estimate values of the model pathways linking the variables, is undertaken below. The squared multiple correlation values of the two models indicate that the predictor variables of the models account for 0.27 (Model 2a) and 0.28 (Model 2b) of the variance in the outcome variable, *Influence of the Use of Program Theory in the Final Report*. Furthermore, the Cronbach's Alpha of the scale measuring the outcome variable *Influence of the Use of Program Theory in the Final Report* indicates the scale reliability is substantial ( $\alpha = 0.81$ ).

A consideration of the mean score for each evaluation gives some insight into the differences between the impact of the evaluation findings, as measured by the *Influence of the Use of Program Theory in the Final Report* outcome variable. The mean scores for the Influence of the Use of Program Theory in the Final Report scale, which provides the measure for the *Influence of the Use of Program Theory in the Final Report* outcome variable, indicate the Pre-Apprenticeship Program stakeholders (mean score 6.0) are the least influenced by the use of program theory in the final report, while the MCJ program stakeholders (7.2) were most influenced, and the YOHFest Program stakeholders (mean score 6.2) fell between the two.

*Program Causative Theory, Model 2a.* The direct effect and regression estimate values indicate that the link between the *Program Causative Theory* latent variable and the outcome variable, *Influence of Use of Program Theory in the Final Report*, is an insignificant one (Direct Effect= 0.08). However, the links between *Program Causative Theory* and *Evaluation Study Characteristics*, *Evaluation Study Characteristics* and *Commitment to the Study (Post)*, and *Commitment to the Study (Post)* through to the outcome variable all have significant regression values ( $p \leq 0.10$ ). The regression estimate value indicates the small negative link between *Program Causative Theory* and the *Evaluation Study Characteristics* to be substantial ( $p \geq 0.10$ ).

As mentioned above in the discussion of the findings regarding Model 1a and Model 1b, a review of the qualitative responses of subjects enables further insight into possible reasons for the lack of impact of the use of *Program Causative Theory* in the undertaking of the evaluation on the impact of the evaluation information. As the reasons are similar to those identified above they are not detailed again here. Nevertheless, it may be that had the evaluation process been more optimal, a greater effect of the use of program causative theory on the evaluation's information may have occurred.

Generally, the qualitative information gathered from stakeholders of the evaluation study, which was classified highest on causative theory use in the undertaking of the evaluation study, indicates that although the stakeholders were happy with the use of program theory in the undertaking of the evaluation, they were unhappy with other characteristics of the evaluation study, the implications of which will be discussed in Chapter IX. As only two of the evaluations included as cases for the study were classified as having used program causative theory in undertaking of the evaluation, the results of the study may have been skewed by the responses of one group. Therefore, the regression values seem to indicate that the use of program causative theory in the evaluation has had a small negative impact on stakeholders' perceptions of evaluation study characteristics. The qualitative information, however, appears to indicate that the stakeholders' perception of the evaluation study



characteristics is a reflection of the evaluation study process, not the use of program causative theory, *per se*.

*Program Implementation Theory, Model 2b.* The regression estimate value indicates that the *Program Implementation Theory* variable is not a predictor of the outcome variable, *Use of Program Theory in the Final Report*. These findings seem to indicate the use of program implementation theory in undertaking of an evaluation study has no effect on the use of program theory in the final report. These results are surprising as it was expected that the greater the use of program implementation theory in the evaluation process, the more detail would be included in the final report specific to the implementation theory of the program, thereby increasing the likelihood of the report having an effect on the thinking and practice of stakeholders with regards to the program, and any decisions they have made, or will make, regarding the program. This finding does not support the arguments of numerous authors suggesting that the use of the program's implementation theory focuses an evaluation study on issues pertinent to why the program was, or was not, delivered as intended, enabling valuable insights into the program delivery and thereby yielding evaluation information more likely to be put to use (Chen, 1989; Funnell, 1997; Scheirer, 1987; Scheirer and Rezmovic, 1983).

As detailed earlier in this chapter regarding the *Influence of the Study Findings* outcome variable, each of the evaluation studies classified with use of program implementation theory, included here as a case for investigation, have issues which are likely to have an impact on the effect of the studies and their information. For instance, MCJ Program stakeholders employed within the Curriculum Directorate of the Department of Education and Training (DET) central office had a very good understanding of both the implementation program theory and causative program theory underpinning the program and therefore they were already quite familiar with the program theory detailed in the final report. They also had clear avenues through which to put the evaluation information to use. However, the next stage of the MCJ Program was already in motion when the final report of the evaluation study became

available. Hence, a significant opportunity for the evaluation's information to have an impact had passed. In addition, a number of the MCJ Program stakeholders were not in a position to put the information of the evaluation to use in either decisions or practice regarding the program, as they were located at a distance from the central DET office, though the evaluation did have some influence on their thinking.

In general, the stakeholders of the Pre-Apprenticeship Program, although very keen to get information at the outset of the evaluation study regarding program implementation, indicated in responding to relevant items in the Stakeholder II Interview that, in fact, they found the program theory information included in the final report to be interesting, but not surprising. A few stakeholders felt the value of the report lay in its formalisation of the program implementation logic by putting it in writing. The information itself was not new to them, but having a tangible written form was. In fact, of the three evaluations the stakeholders of the Pre-Apprenticeship Program scored the lowest on the *Influence of the Use of Program Theory in the Final Report* (mean score 6.0).

Although the report had been available for roughly two months prior to the conduct of the Stakeholder II Interviews, the Pre-Apprenticeship Program stakeholder responses indicated that the report had not had an impact on their practice, nor on their decisions, as none had yet been made regarding the program. Generally, it was expected that the evaluation report would be one of multiple sources of information considered in decisions regarding the future of the program. In fact, two stakeholders with considerable influence on the Pre-Apprenticeship Program, who were working within the central office of DET, felt certain that the program would be discontinued as DET was considering alternative avenues for training young people in the various trades areas.

*Evaluation Study Characteristics.* The path from the *Evaluation Study Characteristics* predictor to the outcome variable, *Use of Program Theory in the Final Report*, is not significant. The *Evaluation Study Characteristics* latent variable was estimated to have a small positive total effect on the outcome variable *Influence of the Use of Program Theory in the Final Report*,

in both models 2a and 2b (total effect = 0.28 and 0.2, respectively). The direct effect values of the models indicate only a weak negative effect on the outcome variable (Direct Effect = -0.12 and -0.13, respectively).

A review of the regression estimates indicates the path linking *Evaluation Study Characteristics* and *Commitment to the Study (Post)* and the path linking *Evaluation Study Characteristics* and *Process Use* are both significant ( $p \leq 0.10$ ). The indirect effect value of *Evaluation Study Characteristics* on the outcome variable indicates a stronger positive relationship (indirect effect = 0.41 for both models). Although the regression estimate value of the path from *Process Use* to the outcome variable is not meaningful, the path from *Commitment to the Study (Post)* to the outcome variable is. This matter will be considered further in the section below focusing on these variables. The finding that *Evaluation Study Characteristics* has an effect on the impact of the *Use of Program Theory in the Final Report* confirms the conclusions of numerous authors such as Cousins and Leithwood (1986), Leviton and Hughes (1981), and Siegel and Tuckel (1985).

*Commitment to the Study (Pre and Post)*. The predictor variable *Commitment to the Study (Pre)* has no effect, either direct or indirect, within either model. However, *Commitment to the Study (Post)* on the other hand, has a strong positive effect (direct and total effect = 0.59) on the outcome variable, *Influence of Use of Program Theory in Final Report*, in both Structural Model 2a and Structural Model 2b ( $p \leq 0.10$ ). Furthermore, the regression estimate value is significant ( $p \leq 0.10$ ) for both models. In general, the predictor to outcome estimates of Structural Models 2a and 2b appear to indicate that the greater the commitment of stakeholders, defined as involvement in the evaluation study, to the evaluation after it has been completed, the more likely they are to be influenced by the use of program theory in the report. This finding supports the conclusions of other researchers including Cousins and Leithwood (1987, 1993) and Cummings (1997) that stakeholder commitment to an evaluation positively affects their use, or intention to use, the evaluation information.

A consideration of the paths in the model indicated to be meaningful by the regression estimate values, reveals a link of significant paths in the model from *Program Implementation Theory* and *Program Causative Theory* to *Evaluation Study Characteristics* through to *Commitment to the Study (Post)* and on to the outcome variable, *Influence of Use of Program Theory in Final Report*. This path will be further discussed in the summary of this section.

*Process Use*. There is a weak direct effect of *Process Use* (Direct Effect = 0.15 and 0.16, respectively) on the effect of the use of program theory in the final report; although as indicted by the regression estimate value the link is not important. Therefore, these findings do not confirm the views of Patton (1997, 1998, 2004) that stakeholders who are positively impacted by their involvement in the evaluation process are slightly more likely to be influenced by the evaluation's information.

#### SUMMARY

The significant paths of Structural Model 2a and Structural Model 2b are the same. Both *Program Causative Theory* and *Program Implementation Theory* appear to be important latent predictor variables of the *Evaluation Study Characteristics* variable. Regression estimate *p* values indicate that the *Evaluation Study Characteristics* predictor variable is significantly linked with the *Process Use* variable and the *Commitment to Study (Post)* variable. Only the *Commitment to Study (Post)* has a significant direct link to the outcome variable, *Influence of Use of Program Theory in the Final Report*.

**OUTCOME VARIABLES:**      **INVOLVEMENT IN PROGRAM THEORY  
ELABORATION (STRUCTURAL MODEL (3A & 3B)  
USE OF INFORMATION (STRUCTURAL MODELS  
(4A & 4B)**

The core models, 3a, 3b, 4a and 4b were all found not be a very good fit for the data from which they were estimated, as the *p* values of all models fell below the 0.05 level of significance. Nonetheless, the model diagrams have

been included in the previous chapter and the significant pathways of each model indicated. Due to the poor fit of the models, the information regarding the regression estimates and effect values is likely not reliable. However, it is interesting to note the information from the models is generally supportive of the findings presented above with regards to models 1a, 1b, 2a & 2b. Furthermore, the Cronbach's Alpha of the scale measuring the outcome variable of scales 3a and 3b, *Involvement in Program Theory Elaboration*, indicates that the scale's reliability is very good ( $\alpha = 0.75$ ). The Cronbach's Alpha of all items together of the four scales (*Likelihood of Use 1*, *Likelihood of Use 2*, *Importance of Use 1*, and *Importance of Use 2*) measuring the outcome variable of scales 4a and 4b, *Use of Information* indicates the reliability of the items together is also very good ( $\alpha = 0.91$ ).

#### **CONCLUSIONS REGARDING THE CORE STRUCTURAL EQUATION MODEL OF THE STUDY**

The four structural models which fit the data are core models 1a, 1b, 2a and 2b. These models account for 0.45, 0.46, 0.27 and 0.28 of the variance of the model outcome variables, respectively. Structural models 3a, 3b, 4a, and 4b do not fit the data.

Core Model 1a has only two significant paths as indicated by the regression estimates. These are the paths linking *Evaluation Study Characteristics* to *Commitment to the Study (Post)*, and the *Process Use* variable to the outcome variable, *Influence of Study Findings*. The lack of effect of the *Program Causative Theory* variable in the model is unexpected, as the relevant literature indicates that the use of program causative theory in the undertaking of an evaluation leads to an increased effect of the evaluation and the information it yields. However, the two cases included in this study classified with a higher use of causative program theory, have issues relevant to the stakeholders' perception of *Evaluation Study Characteristics*. Considerations of the qualitative responses of the stakeholders indicate these issues may have limited the effect of the evaluation's information. Unfortunately, the situation is not improved with Core Model 2a in which the *Program Causative Theory*

has a weak negative, yet significant, link to *Evaluation Study Characteristics* that moderates the weak indirect effect of the *Program Causative Theory* variable on the outcome variable, *Influence of the Use of Program Theory in the Final Report*. With regards to Core Model 2a, qualitative responses of stakeholders again indicate issues with the evaluation process specific to each of the programs, with a greater use of program causative theory in undertaking the evaluation study being responsible for the reduced impact of the use of program causative theory.

The *Evaluation Study Characteristics* variable has an important influence on the outcome variable of all models with the exception of Core Model 1a. The path of impact is not obvious with regards to Core Model 1a, in which regression estimates indicate *Evaluation Study Characteristics* only has significant effect on *Commitment to the Study (Post)*. However, *Evaluation Study Characteristics* has significant links with the *Process Use* variable, *Commitment to the Study (Post)* variable and the outcome variable, *Influence of Study Findings* in Core Model 1b. Similarly, in Core Model 2a and Core Model 2b the *Evaluation Study Characteristics* variable has meaningful effect on the *Process Use* variable and the *Commitment to Study (Post)* variable, although the link with the outcome variable is not significant. *Commitment to the Study (Post)*, however, has a significant link with the outcome variable, *Influence of Use of Program Theory in the Final Report*, in both models 2a and 2b, enabling an indirect link between the *Evaluation Study Characteristics* variable and the outcome variable.

A review of the qualitative data collected in the conduct of the study interview gives some insight regarding possible influences to the findings of structural models. With regard to characteristics of the evaluation process, important events in the undertaking of the YOHFest Program evaluation has likely led to a reduced effect of the evaluation and the findings. Although qualitative responses of the stakeholders indicate they found the theory elaboration process very valuable, disillusionment with other evaluation processes has possibly obscured the impact.

The Making Consistent Judgment program evaluation, which included the use of both causative and implementation program theory in the undertaking of the evaluation, also had some specific characteristics which had impact on the effect of the evaluation and the information it yielded. In particular, only stakeholders located in the Curriculum Directorate within the central office for the Department of Education and Training had clear opportunity to put the evaluation's information to use, as the program was centrally driven yet implemented locally. Furthermore, the next round of the MCJ professional development program was underway by the time the final report of the evaluation included as a case in this study, was released. Therefore, an important avenue to put the evaluation's information to use was no longer available. Finally, with regards to the outcome variable of Core Model 2a and Core Model 2b, *Influence of the Use of Program Theory in the Final Report*, primary stakeholders of the MCJ Program evaluation were very cognisant of the program theory, both causative and implementation. Therefore, the theory detailed in the final report was not new to them in any way.

Similarly, with regards to the Pre-Apprenticeship Program evaluation, primary stakeholders responsible for the development and management of the program indicated the main reason for the instigation of the evaluation of the Pre-Apprenticeship Program was to have a formal statement of what was actually taking place in the implementation of the program. Many Pre-Apprenticeship stakeholders indicated they were not surprised by any description of the program's theory in the final report, as it confirmed what they knew to be true. Furthermore, beyond the primary stakeholder group located within the central office of the Department of Education and Training, many stakeholders were only able to obtain a copy of the final Pre-Apprenticeship Program evaluation report by request. These stakeholders indicated the funding and distribution of resources regarding the Pre-Apprenticeship Program was determined at 'head office', and their role was simply to respond. Therefore, their incentive to process the report and acknowledge its contents in their decisions and practice was very limited.

## DEVELOPMENT OF THE FULLY ELABORATED MODEL OF THE STUDY

As was noted in the previous chapter, analysis of the Fully Elaborated Structural Model found it did not fit the data. Therefore, development of the Structural Model involved the addition of one latent variable of the Elaborated Model at a time, to all eight Core Models. The latent predictor variables added in the Elaborated Model and not included in the Core Model are; *Organisational Learning Environment*, *Evaluator Characteristics*, *Stakeholder Characteristics*, *Expected Study Involvement* and *Study Involvement*. The influence of these variables in the model is further detailed below.

### ORGANISATIONAL LEARNING ENVIRONMENT

Core models 1a and 1b (Outcome Variable: *Influence of Study Findings*, refer Figure 7.12 and Figure 7.13) and models 2a and 2b (Outcome Variable: *Influence of Use of Program Theory in Final Report*), developed with the *Organisational Environment Characteristics* variable were a reasonable fit of the data, as indicated by the  $p$  values (all  $p$  values  $> 0.05$ ) of all four models. Interestingly, the squared multiple correlation values indicate the predictor variable of Model 1a elaborated with the *Organisational Environment Characteristics* variable, account for 0.46 of the variance in the outcome variable, *Influence of Study Findings*, while the predictor variables of Model 1b account for 0.46 of the variance in the outcome variable. The predictor variables of models 2a and 2b, elaborated with the *Organisational Environment Characteristics* variable, both account for 28% of the variance in the outcome variable, *Influence of the Use of Program Theory in the Final Report*.

*Organisational Learning Environment*. With regards to the models with *Influence of Study Findings* as an outcome variable, *Organisational Learning Environment* was indicated by the regression estimate values and direct effect values (Direct Effect = -0.22) to have a small negative impact on the *Evaluation Study Characteristics* variable, though no effect on the outcome



variable (Direct Effect = 0.01). This finding indicates the greater the stakeholder's perception of the extent to which the environment of the organisation has the characteristic of a learning environment, the lower their perception of the *Evaluation Study Characteristics*. It is strongly suspected this result is a reflection of the effect of two factors. First, due to the way the data has been managed (*Organisational Environment Characteristics* interview subjects are different to the Stakeholder II Interview subject group) the value of each subject regarding the *Organisational Environment Characteristics* observed variable is the mean of the subject aggregate value for each program to the five *Organisational Environment Characteristics* indicator variables. Second, two programs have a slightly lower score (MCJ 153.3, Pre-Apprenticeship 153.2, YOHFest 207.75) while the third is slightly higher. In fact, a review of the relevant qualitative data strengthens the indication that the YOHFest organisation actually presents the characteristics considered to be indicative of a Learning Organisation. Third, as detailed previously in this chapter, YOHFest Program stakeholders had some issues specific to the evaluation process of the YOHFest Program which, it is suspected, has led to their lower response scores with regards to the observed variables measuring the *Evaluation Study Characteristics* and the outcome variables. Therefore, it seems the slight negative effect of the *Organisational Environment Characteristics* variable on the *Evaluation Study Characteristics* variable, strongly reflects disposition of the YOHFest Program evaluation. Had more cases been included in the investigation, it is likely the issues specific to the YOHFest Program evaluation would have had less impact.

*Program Theory.* Neither *Program Causative Theory* nor *Program Implementation Theory* variables have a meaningful effect on the outcome variables in the development of models 1a, 1b, 2a, or 2b with the *Organisational Learning Environment* variable added. However, both *Program Implementation Theory* and *Program Causative Theory* appear to be significant predictors of *Evaluation Study Characteristics*, as indicated by the regression estimate values of the path in all four models.

*Evaluation Study Characteristics.* The *Evaluation Study Characteristics* variable has a significant link with the *Process Use* variable, the *Commitment to Study (Post)* variable, and the outcome variable, *Influence of Study Finding*, in the development of models 1a and 1b. In terms of effect values, the effect of the *Evaluation Study Characteristics* variable on the *Influence of Study Findings* variable is a very strong positive one, underpinned by a strong positive direct effect and a weak positive indirect effect.

In the development of models 2a and 2b, the link between the *Evaluation Study Characteristics* variable and the outcome variable, effect of *Influence of Program Theory in the Final Report* is not significant. However, all other significant paths from the *Evaluation Study Characteristics* variable are the same as those for the development of models 1a and 1b. The nature of the relationship is slightly different in developed models 2a and 2b, however. In these models, developed with the *Organisational Learning Characteristics* variable, the primary effect of the *Evaluation Study Characteristics* is a stronger positive indirect impact, while the direct effect value (-0.17) indicates the *Evaluation Study Characteristics* variable has a weak negative impact on the outcome variable, *Influence of the Use of Program Theory in the Final Report*. These findings are in agreement with those of Model 2b.

*Process Use.* In developed models 1a and 1b the *Process Use* variable has a small to strong significant effect on the outcome variable, *Influence of Study Findings*, as indicated by the effect values and regression estimates (Direct Effect = 0.30). However, with regards to the developed models 2a and 2b, the effect of *Process Use* on the outcome variable is not significant. These findings are in agreement with those of the testing of models 2a and 2b detailed in the previous section.

*Commitment to the Study (Pre & Post).* *Commitment to the Study (Pre)* has no meaningful impact, either direct or indirect, in any of the four models developed with *Organisational Learning Characteristics*. *Commitment to the Study (Post)*, However, although not a significant effect on the outcome variable in developments of models 1a and 1b, has a strong positive significant

impact on the outcome variable in developments of models 2a and 2b, as indicated by the regression estimates and effect values (Total Effect = 0.58). Therefore, it seems the stakeholder's commitment to the evaluation study measured as involvement, has effect on the extent to which the *Use of Program Theory in the Final Report* has an effect on them, but does not have a meaningful impact on the extent to which they are influenced by the findings of the study, when the predictor variables of the model include *Organisational Learning Environment*. The findings are in agreement with those of the testing of models 1a, 1b, 2a and 2b, detailed in the previous section of this chapter.

#### SUMMARY

The findings from the testing of models 1a, 1b, 2a and 2b developed with the *Organisational Environment Characteristics* variable are generally the same as those from the testing of the core models 1a, 1b, 2a and 2b. The *Organisational Environment Characteristics* variable has a small negative influence on the *Evaluation Characteristics* latent variable, though no significant effect on the outcome variable. It is suspected that the negative nature of the regression estimate value is a reflection of three factors. These three factors being: the manner in which the *Organisational Environment Characteristics* data was managed (e.g. one value for subjects of each program); the fact that the YOHFest Program has the highest *Organisational Environment Characteristics* value, while the other two representative organisations scored roughly the same with regards to the organisational environment; and the issues with the YOHFest Program evaluation process which likely detracted from greater use of the evaluation and the information it yields. Regression estimate values indicate all significant links between other predictor variables of the core model and between predictor variables and the outcome variables are the same as in the testing of core model 1a and 1b. Therefore, in terms of bringing in relevant qualitative information to further express the findings, nothing can be added to that which has already been given in the previous section detailing the findings in the testing of the core models.

## EVALUATOR CHARACTERISTICS

Core models 1a and 1b (Outcome Variable: *Influence of Study Findings*, refer Figure 7.12 and Figure 7.13) and models 2a and 2b (Outcome Variable: *Influence of Use of Program Theory in Final Report*), developed with the *Evaluator Characteristics* variable, were a reasonable fit of the data, as indicated by the  $p$  values (all  $p$  values  $> 0.05$ ) of all four models. Interestingly, the squared multiple correlation values indicate the predictor variable of Model 1a elaborated with the *Evaluator Characteristics* variable account for 0.46 of the variance in the outcome variable, *Influence of Study Findings*, while the predictor variables of Model 1b account for 0.46 of the variance in the outcome variable. The predictor variables of models 2a and 2b elaborated with the *Evaluator Characteristics* variable, both account for 29 % of the variance in the outcome variable, *Influence of the Use of Program Theory in the Final Report*.

*Evaluator Characteristics.* The regression estimate values of models 1a, 1b, 2a, and 2b indicate the *Evaluator Characteristics* variable has a small negative, yet significant, effect on the *Evaluation Study Characteristics* latent variable in all four models, though no significant impact on the outcome variable. It is suspected this finding is primarily a result of the YOHFest Program evaluation. The value for the *Evaluator Characteristics* observed variable is the mean of the subject aggregate value for each evaluation to the four observed variables measuring *Evaluator Characteristic* (*Evaluator Practice Scale 1*, *Evaluator Practice Scale 2*, *Evaluator Practice Scale 3*, *Evaluator Perception of Stakeholder Involvement in the Evaluation*) indicator variables. Two program evaluations have a slightly lower score (MCJ 68.50, Pre-Apprenticeship 63.50 YOHFest 73.50), while the third is slightly higher. In fact, a review of the relevant qualitative data strengthens the view that the YOHFest program evaluators actually indicate they value the characteristics considered to be indicative of evaluators interested in involving stakeholders in the evaluation and in adopting a partnership-like relationship with evaluation stakeholders in the conduct of the study. As discussed previously in this chapter, most of the YOHFest Program stakeholders were unhappy with

some aspects of the evaluation conduct and report dissemination. YOHFest Program evaluators, however, had the highest values on the observed variables providing the measures for the observed *Evaluator Characteristics* variable, indicating they were, of the three evaluator groups, the most open to stakeholder involvement in the evaluation process, and more open to adopting an evaluation approach in which they worked in partnership with the evaluation stakeholders. Unfortunately, circumstances resulted in some incongruencies between stakeholder expectations, regarding the quantitative data they expected to be collected and reported in the evaluation, and what was actually collected. The regression estimates of the four models indicate the *Evaluator Characteristics* latent variable has no significant effect on the outcome variable.

*Program Causative Theory and Program Implementation Theory.* Similar to the models elaborated with the *Organisational Environment Characteristics* variable, models 1a, 1b, 2a and 2b indicate the *Program Causative Theory* variable has a significant small negative effect on the outcome variables, while the *Program Implementation Theory* has a small significant positive impact on the outcome variables. Possible reasons for these values have been discussed with regards to previous models. The variables do not have a significant impact on the latent outcome variables in any of the four model elaborations.

*Evaluation Study Characteristics.* The regression estimate values of the models indicate that the *Evaluation Study Characteristics* variable has a significant effect on both the *Process Use* and *Commitment to Study (Post)* variables, indicating that the more positively stakeholders of the evaluation perceive the characteristics of the evaluation study, the more likely they are to be committed to the study and be effected by their involvement in the study. No regression estimate value indicates *Evaluation Study Characteristics* to be a significant predictor of either of the two outcome variables.

*Process Use.* Although regression estimate values indicate the *Process Use* variable is not a significant predictor of the latent outcome variable *Use of Program Theory in the Final Report* in either model 2a or 2b elaborations, it is

a variable indicated to be a small significant predictor of the outcome variable *Influence of Study Findings* (models 1a and 1b). These findings are consistent with the findings of the models elaborated with *Organisational Environment Characteristics* variables, and those of models 2a and 2b.

*Commitment to Study (Pre)* and *Commitment to Study (Post)*. Although *Commitment to Study (Pre)* has no significant influence on another variable of the model, *Commitment to Study (Post)* has significant influence on the outcome variable in models 2a and 2b. This link is not significant in models 1a and 1b.

#### SUMMARY

The findings from the testing of models 1a, 1b, 2a and 2b developed with the *Evaluator Characteristics* variable, are generally the same as those from the testing of core models 1a, 1b, 2a and 2b. The *Evaluator Characteristics* variable has a small negative influence on the *Evaluation Characteristics* latent variable, though no significant effect on the outcome variable. It is suspected that the negative nature of the regression estimate value is a reflection of three factors. These three factors being: the manner in which the *Evaluator Characteristics* data was managed (e.g. one value for subjects of each program); the YOHFest Program having the highest *Evaluator Characteristics* value while the other two representative organisations scored roughly the same with regards to the *Characteristics of the Evaluator*; and the issues with the YOHFest Program evaluation process which likely detracted from greater use of the evaluation and the information it yields. Regression estimate values indicate all significant links between other predictor variables of the Model and between predictor variables and the outcome variables are the same as in the testing of core models 1a and 1b. Therefore, in terms of bringing in relevant qualitative information to further express the findings, nothing can be added to that which has already been given in the discussion of the findings of the previous models.

## INVOLVEMENT (EXPECTED AND POST)

The fit statistics of the eight models developed with *Expected Study Involvement Variable* and the *Study Involvement (Post)* variable were all found to be a poor fit of the data. The developments of core models 1a and 1b, however, were not far from fitting ( $p = 0.04$ ). Therefore, the findings of these two models have been taken into consideration and are detailed in this section. Interestingly, the squared multiple correlation values of core models 1a and 1b, developed with the *Study Involvement (Expected)* and *Study Involvement (Post)* variable, indicate the predictor variables of the models account for 0.50 and 0.51 of the variances in the outcome variable, *Influence of Study Findings*, respectively.

*Study Involvement (Expected and Post)*. The regression estimate information and effect information indicate *Study Involvement (Pre)* has a significant positive, yet weak, effect on the *Stakeholder Involvement (Post)* variable in the elaborations of both Model 1a and Model 1b. This finding is in line with the logic underpinning the inclusion of the two variables. The expectation was that stakeholders with a greater expectation of being involved in the evaluation study undertaking would be more likely to actually be involved in the study. Furthermore, the *Study Involvement (Post)* variable, as indicated by regression estimates of the models and effect information, has a significant small positive direct impact on the outcome variable, *Influence of Study Findings*. This finding supports the main conclusion of Cumming (1997), from which the items focusing on involvement were taken. He found involvement to have substantial effect on the extent to which stakeholders used the information of an evaluation study.

*Program Theory*. The *Program Causative Theory* variable had significant impact on the *Evaluation Study Characteristics* variable, yet not on the outcome variable, *Influence of Study Findings*. The impact of the *Program Implementation Theory* variable is the same, having a meaningful effect on the *Evaluation Study Characteristics* variable, yet not on the outcome variable, *Influence of Study Findings*.

*Evaluation Study Characteristics.* With regards to the elaborations of models 1a and 1b, the relevant regression estimates indicate that the *Evaluation Study Characteristics* variable does have a significant effect on *Commitment to the Study (Post)*, *Process Use* and *Influence of Study Findings* outcome variable. The effect of the *Evaluation Study Characteristics* variable on the outcome variable is a very strong positive one (Total Effect = 0.57); with both the direct and indirect paths to the outcome variable having a small positive impact (Direct Effect = 0.33, Indirect Effect = 0.24).

*Process Use.* The regression estimate values and effect values (Direct Effect = 0.23) indicate that the *Process Use* variable has a small positive significant influence on the outcome variable, *Influence of Study Findings* in both of the core models (1a and 1b) developed with the *Study Involvement* variable. These findings seem to confirm the logic of Patton (1997, 1998, 2004) that process use of the evaluation information, or use of evaluation information gained through involvement in the evaluation process, is likely to have an effect on the extent to which stakeholders are affected by the study findings following the conclusion of the evaluation study.

*Commitment to the Study (Pre and Post).* *Commitment to the Study (Pre)* has no significant relationship, either direct or indirect, with the outcome variable, *Influence of Study Findings*. Effect values indicate *Commitment to the Study (Post)* has a weak positive impact on the *Influence of the Study Findings* (0.16 Total Effect). The regression estimate value, however, indicates the relationship is not significant.

## SUMMARY

The inclusion of the *Study Involvement* latent predictor variables (Expected and Post) increases the amount of variance in the outcome variable, *Influence of Study Findings*, accounted for in the model. The greatest amount of variance accounted for in the outcome variable of any other model considered in this chapter is the Core Model 1a ( $r^2 = 0.45$ ). Core models 1a and 1b with the *Study Involvement* predictor variables included, increases to 50% and 51%,



respectively. In addition, the *Expected Study Involvement* variable has significant effect on the *Study Involvement (Post)* variable, which in turn has a significant direct effect on the outcome variable.

## CONCLUSION

Analyses of the core structural equation models of this study have yielded some interesting information. Four models were found to fit the data, two with the outcome variable, *Influence of Study Findings*, and two with the outcome variable, *Influence of the Use of Program Theory in the Final Report*. With regards to the primary purpose of this study, which is to investigate the impact of the uses of program theory in the undertaking of the evaluation, the results are slightly different to those suggested by the relevant work of other authors.

In the development of the Evaluation Classification Matrix in Chapter III of this dissertation, it was found that theory-based evaluation studies tend to lean primarily towards the inclusion of one type of program theory or the other (e.g. *Program Causative Theory* over *Program Implementation Theory*) in the undertaking of the evaluation study. A Pearson Bivariate correlation analysis of the data from the classification of the 21 studies included in the Classification Matrix of Chapter III and the three cases investigated for this study (24 cases in total) indicates the use of *Program Implementation Theory* and *Program Causative Theory* in the undertaking of an evaluation study is negatively correlated (-0.15), though the correlation is not significant (0.49) at the 0.05 level. Considering this information, a decision was made to consider the two variables, *Program Implementation Theory* and *Program Causative Theory* separately, rather than as two halves of the same variable. The results of the study seem to indicate that the separate consideration of the two is a reasonable option, as they seem to have differing effects on the evaluation study process and the effect of the evaluation information. This is discussed further discussed below.

One reason for the possible disparity between the two types of theory use is due to their nature. Program implementation theory is the more tangible of the

two theory types, specifically operationalised in the activities of the program and the effects, or outcomes, of each activity measurable in concrete terms. Program causative theory is less tangible and difficult to measure as it is often only realised in shifts in subjects' attitudes. Measures of changes in the knowledge of subjects, or changes in behaviour, may only be linked implicitly to attitudinal shifts. Furthermore, changes to elements of the program implementation theory of a program are much more easily managed than the program causative theory upon which a program is based. In particular, the effect of program causative theory in the testing of the core models as indicated by the regression estimate and effect values of models 1a and 2a is of interest. In the first model, the *Program Causative Theory* variable was found to have no significant direct or indirect effect. However, in the second it actually appears to have a significant negative effect on the stakeholder's perceptions of the *Evaluation Study Characteristics*. Relevant literature, in fact, suggests the opposite to be true: the greater the use of program causative theory in the undertaking of the evaluation, the higher regard for the evaluation process and the impact of the study findings. As detailed in the relevant sections of this chapter, it is suspected these findings are a result of issues of the evaluation studies included as cases. In particular, the YOHFest Program evaluation, which was classified as the highest of the three cases in the use of program causative theory in the undertaking of the evaluation included in this study, had some problems in the evaluation study process which led stakeholders to view the evaluation study with some negativity and likely reduced the use of the evaluation study and its findings. As only three studies have been included for investigation in this study, the findings of one case have enough weight to significantly impact the findings of the overall investigation.

The results of the structural equation modelling process seem to indicate the effect of the *Program Implementation Theory* latent variable in the undertaking of the evaluation has a positive significant effect on the *Evaluation Study Characteristics* latent variable and on the outcome variable indirectly. Although the finding of a stronger direct impact of the use of program implementation theory on the outcome variables would have been in

line with those suggested by other authors, the findings of this study are not entirely at odds. As detailed above, it is expected that events and situations specific to the evaluations included as cases in this investigation have limited the direct effect of program implementation theory variable. Once again, due to the limited number of cases included in this study, each one has avenue to have meaningful effect on the overall findings of the study.

## CHAPTER IX

### SUMMARY AND CONCLUSIONS

The main purpose of this study was to investigate the singular and interactive effect of ten factors on the utilisation of evaluation information as measured by the four outcome variables. The influence of program theory was of primary interest. Of the ten factors, five primary factors which were expected to affect the use of evaluation information were included in the Core Model of the study. Eight versions of the Core Model were developed and tested, one pair (one including *Program Causative Theory* and one including *Program Implementation Theory*) for each of the four outcome variables. A study of the full model with the ten factors was also undertaken using a process of model development.

This chapter will first provide a summary of the investigation, including a brief synopsis of the strengths and weaknesses of the study. The main findings and the conclusions which have been drawn from them are then considered. Finally, the implications of the results of this investigation for further research and for evaluation practice are considered.

### A SYNOPSIS OF THE STUDY

A review of the pertinent literature indicated that there was a need to expand what is known about the use of the information from evaluation studies, and the factors which impact on that use. Of particular interest to this investigation is the gap identified in the literature regarding the influence of program theory on the use of evaluation information. Although considerable literature examines the benefits of a theory-based evaluation on the influence and use of the evaluation information, no empirical studies were found which had investigated the link. This study follows on from the work of a number of authors concerned with investigating the predictors of evaluation information use (Alkin et al., 1979; Cousins and Leithwood, 1986, 1993; Kirkhart, 2000; Leviton and Hughes, 1981; Seigel and Tuckel, 1985; Shadish, Cook and

Leviton, 1991). The logic underpinning this work was that if evaluators were cognisant of the factors which shape the potential of an evaluation's impact, then these factors could be incorporated into the design and implementation of an evaluation study.

In consideration of the factors indicated by previous research to have an effect on the use of an evaluation's information, a conceptual model incorporating program theory and other factors expected to have the most effect on evaluation information use, was developed. The logic of this conceptual model underpinned the design, methods, instrument development, data analysis and structural model development of this study, giving rise to the following research questions:

- i. What is the influence of program theory on the use of evaluation information?
- ii. Which factors have the greatest impact on the use of evaluation information?
- iii. How do these factors interact with each other to affect use?

Three empirical case studies of concurrent program evaluation studies were undertaken with a view to identifying the use of the information of each evaluation, and determining the characteristics of each case. The longitudinal study design adopted both quantitative and qualitative research methods in the investigation of each case. Six interview schedules were developed for data collection and were used in interviewing a total of 64 stakeholders and evaluators. Each of these interview schedules contain a mix of open response and closed response items. The information gathered from the open response items was collected to add insight to the quantitative data. Twenty four scales are contained within the six interview schedules. Following a reliability analysis of the 24 scales, further analysis of the data was undertaken through Structural Equation Modelling (SEM) using the Analysis of Moment Structures, or AMOS version 7.0 structural equation modelling software package.

The scales are the measures for the observed variables providing the values for the ten latent predictor variables and the four outcome variables of the Fully

Elaborated Model of the study, described in Chapter III. The nine latent predictor variables of the model are: *Organisational Learning Environment Characteristics*; *Evaluator Characteristics*; *Stakeholder Characteristics*; *Program Theory (Program Implementation Theory and Program Causative Theory)*; *Commitment to the Study (Pre)*; *Commitment to Study (Post)*; *Expected Study Involvement*; *Study Involvement (Post)*; and *Process Use*. The four different outcome variables are: *Influence of Use of Program Theory in the Final Report*; *Influence of Involvement in Program Theory Elaboration*; *Use of Evaluation Information*; and the *Influence of Study Findings*. Each of the outcomes was used separately as they are conceptually different. This strategy, more fully detailed in Chapter VII, was adopted to facilitate model fit in view of the small sample size of the study. Furthermore, it was expected that the influence of the predictor variables on the outcome variables would be clearer in simpler models.

The analysis of the Fully Elaborated Model determined that it did not fit the data from the three case studies. A Core Model was developed which included a common set of five latent variables predicting, in turn, the four outcome variables. The five Core Model latent predictor variables are: *Program Theory (Causative Program Theory and Implementation Program Theory)*; *Evaluation Study Characteristics*; *Process Use*; *Commitment to Study (Pre)*; and *Commitment to Study (Post)*. Eight separate versions of the Core Model were developed and tested, one pair for each outcome variable. In one version of each of the pairs, the program theory variable is represented by *Program Implementation Theory*, while the other version includes the *Program Causative Theory* variable. Each model was tested and the results reported in Chapter VII. Core model versions 1a and 1b (Outcome Variable: *Influence of Study Findings*) yielded the strongest fit statistics and accounted for 0.45% and 0.46% of the variance in the outcome variable of the Model, respectively. Core Model versions 2a and 2b (Outcome Variable: *Influence of the Use of Program Theory in the Final Report*) are also an acceptable fit of the data and accounted for 0.27% and 0.28% of the variance in the outcome variable of the model, respectively.

A model development process testing for model fit was then undertaken in which the Core Model was elaborated by testing five modified models. Four models were modified and tested by adding a single predictor variable to the model. Model versions 1a, 1b, 2a, and 2b, elaborated with *Organisational Environment Characteristics* and with *Evaluator Characteristics* latent variables, all provide an acceptable fit of the data. The addition of one predictor variable did not significantly influence the amount of variance accounted for in the outcome variable. The fifth model was modified by adding two latent predictor variables. Model versions 1a and 1b elaborated with *Expected Study Involvement* and *Study Involvement (Post)* latent variables, although only a close fit with the data, accounted for 51% per cent of the variance in the outcome variable *Influence of Study Findings*.

#### CONCLUSIONS OF THE STUDY

The three research questions which have been the focus of this study are used to summarise the conclusions of the study. Only model versions with the outcome variables *Influence of Study Findings* and *Influence of the Use of Program Theory in the Final Report* were found to fit the data at an acceptable level. Therefore, consideration of the factors found to have the most effect will centre on model versions with these two outcome variables. Model versions with *Influence of Involvement in Program Theory Elaboration*, and the *Use of Evaluation Information* as the outcome variables were not found to have an acceptable fit with the data.

#### ***What is the influence of program theory on the use of evaluation information?***

*Program Causative Theory*. Regression estimates of Model 1a indicate the *Program Causative Theory* variable has no significant direct effect on the outcome variable, *Influence of Study Findings*. Similarly, regression estimate values indicate the *Program Causative Theory* variable is not a significant predictor of the outcome variable, *Influence of Use of Program Theory in the Final Report*, in Model 2a. Finally, in the elaborations of the Core Model with

the *Organisational Characteristics* variable and the *Study Involvement (Expected and Post)* variables, the *Program Causative Theory* variable has no significant effect on the outcome variable.

Given the considerable literature described in Chapter V on the benefits of including a program theory in an evaluation study, the lack of direct effect of the *Program Causative Theory* latent variable on the outcome variables of the models is surprising. A consideration of the qualitative information gathered from stakeholders gives some insight into the findings. The stakeholders of the YOHFest Program evaluation; the evaluation classified as highest on the use of *Program Causative Theory*, were very unhappy with some of the characteristics of the evaluation study. They were unhappy that the final report did not contain the quantitative data they required to support their request for further funding of the program. They also had concerns regarding some of the data collection methods employed in the evaluation and were disheartened that the final report was delayed by nearly a year from the expected date of delivery. They all were, however, very happy with the program theory elaboration process conducted by the evaluators, which occurred early in the evaluation. When first interviewed, all YOHFest Program stakeholders involved in the program theory elaboration process were extremely positive regarding the gains they had made in terms of program insight by being involved in the process. Unfortunately, it seems other characteristics of the evaluation have limited the use of the YOHFest Program evaluation information. Therefore, the findings of this study regarding the impact of *Program Causative Theory* latent variable in the model versions may be misleading, and should be considered in the light of the broader context.

Similarly, stakeholders of the Making Consistent Judgments Program evaluation, which had the second highest value for *Program Causative Theory* use, were very happy with the use of program causative theory in the evaluation. However, issues regarding the nature of program implementation (the program was centrally driven from the head office of the Department of Education and Training) made actual use of the evaluation information difficult for those located in the district offices and schools. Furthermore,



because the program was so strongly driven from the central office there was some disillusionment among stakeholders located in the district offices and schools that their opinions had little weight with regards to both the evaluation and the program. Therefore, again the context of the program and the evaluation, and not necessarily the use of program causative theory on the evaluation, have likely limited the effect of program theory use on the outcome variables. The third study, the Pre-Apprenticeship Program evaluation, was not high on the use of *Program Causative Theory*, but was high on the use of *Program Implementation Theory* and will be discussed.

*Program Implementation Theory.* Regression estimate values indicate that the *Program Implementation Theory* latent predictor variable has no significant effect on any outcome variables. Qualitative data gathered from stakeholders of the evaluation with the highest value regarding program implementation theory use; the Pre-Apprenticeship Program evaluation, indicate use of the evaluation information may have been limited, as the Department of Education and Training was considering alternative avenues for training young people in Western Australia and there was likelihood that the Pre-Apprenticeship Program would be discontinued. There also was some concern among stakeholders regarding the statistics included in the final report of the evaluation which led them to be cautious about relying on the information. Therefore, although the use of program implementation theory appears to positively influence the stakeholders' perception of the evaluation characteristics, other circumstances and factors may have limited the effect of *Program Implementation Theory* on the outcome variables.

As the present study only includes three cases, the particular characteristics of each case has had an opportunity to sway the findings of the study overall. Generally, it seems the context of each case has had an important impact on the effect of program theory on the use of the evaluation's information.

***Which factors have the greatest impact on the use of evaluation information?***

The *Evaluation Study Characteristics* predictor variable appears to be the strongest predictor variable of the models *with Influence of Study Findings* as the outcome variable (models 1a and 1b) and with *Influence of the Use of Program Theory in the Final Report* as the outcome variable (models 2a and 2b). The *Evaluation Study Characteristics* variable is also a strong direct predictor of the outcome variable in model version 1b and model versions 2a and 2b, elaborated with the *Organisational Environment Characteristics* variable, indicating that stakeholder's with a more positive perception of the *Evaluation Study Characteristics* are more likely to be influenced by the study findings when *Program Implementation Theory* use is higher, and when the organisation responsible for the delivery of the program has an environment with characteristics facilitating learning. Also, in all model versions except 1a, the *Evaluation Study Characteristics* variable has indirect influence on the outcome variable of the model, either through *Process Use* or *Commitment to the Study (Post)*. In all instances, the strength of this indirect effect is strong.

The *Influence of Study Findings* variable is significantly affected by the *Process Use* latent variable in core models 1a and 1b, indicating that involvement in the evaluation process has a positive influence on the use of the study findings. *Process Use* does not have a significant impact on the outcome variable *Influence of the Use of Program Theory in the Final Report*. Possibly those stakeholders involved in the evaluation process pre-empt influence of the theory used in the final report through exposure to the program theory logic and description.

The link between *Commitment to Study (Post)* and the outcome variable is only significant when the outcome variable is *Influence of the Use of Program Theory in the Final Report*, with the exception of the model version developed with the *Evaluator Characteristics* latent variable. When the outcome variable is *Influence of Study Findings* (model versions 1a and 1b) the link is not significant in any model development. It is suspected that those committed to

the evaluation are more likely to positively view the final report and are more open to be influenced by its content, particularly if they have a positive perception of the evaluation study characteristics.

Finally, in the model development process undertaken, the only additional variables which had a significant impact on the outcome variable *Influence of Study Findings*, were those included in the model elaboration with the *Expected Study Involvement* and *Study Involvement (Post)* variables, indicating the *Study Involvement* variables may together have a significant positive influence on the *Influence of Study Findings* outcome variable. This influence indicates those who expect to be involved in the study are more likely to be involved and are more likely to be affected by the evaluation information as a result of their involvement. Furthermore, those with greater involvement in the evaluation study are more likely to be influenced by the study findings. This influence should be explored in future research.

#### ***How do these factors interact with each other to affect use?***

Regression estimates of Model 1a indicate the *Program Causative Theory* variable has no significant direct effect on any other variable. However, the regression estimate values of Model 2a indicate the *Program Causative Theory* predictor variable has a small negative effect on the *Evaluation Study Characteristics* variable. Similarly, in the elaborations of the Core Model with the *Organisational Characteristics* variable and the *Study Involvement (Expected and Post)* variables, the *Program Causative Theory* variable has a significant, small yet negative effect on *Evaluation Study Characteristics*. Furthermore, regression estimate values indicate that the *Program Implementation Theory* latent predictor variable has a significant, but small, positive effect on *Evaluation Study Characteristics* latent variable in all models of the study found to fit the data. The negative effect of the *Program Causative Theory* variable on the *Evaluation Study Characteristics* variable is also concerning if confirmed in future studies. As discussed in the above sections, caution must be taken in consideration of the quantitative data.

Thoughtful reflection on the qualitative data, as outlined above, gives insight into the broader context of influences on the evaluation process.

Model 1a indicates a significant link between the *Evaluation Study Characteristics* latent variable and the *Commitment to Study (Post)* latent variable, the *Process Use* variable and the *Influence of Study Findings*. In most of the model versions which fit the data, the link between the *Evaluation Study Characteristics* latent variable and the *Commitment to Study (Post)* latent variable is significant, indicating that the more positively stakeholders of an evaluation view the characteristics of the evaluation, the more committed they are to the evaluation.

Finally, Model 1b in the model versions, elaborated with the *Organisational Environment Characteristics* latent variable, the *Evaluator Characteristics* latent variable and the two *Study Involvement (Expected and Actual)* variables, has a significant link from the *Evaluation Study Characteristics* to the *Process Use* variable and from the *Process Use* variable to the outcome variable. This seems to indicate that those who have been involved in the evaluation process are more likely to view the evaluation study characteristics more positively and, in turn, be influenced by the evaluation findings.

#### STRENGTHS AND WEAKNESSES OF THE STUDY DESIGN

The concurrent longitudinal design adopted in this investigation of the three evaluations had particular merit. It is likely that a retrospective approach to the collection of the study data may have compromised the quality of the data. Stakeholders tend to move with the progress of the program, so asking them to recall details of, for instance, their involvement in the program theory elaboration process, is likely to be subject to both memory loss and changed perceptions due to learning during the study. Furthermore, the concept of program theory often is a difficult one for people to grasp due to its intangible nature. Certainly, asking them to recall details of how their cognitions regarding program theory influenced their thinking about the program, and the nuances of that influence, three months or more after the event, runs the risk of

compromising the quality of the data collected. In addition, the concurrent longitudinal approach of this study enabled the collection of data from sources at three distinct stages of the evaluation process, without the responses of stakeholders being influenced by subsequent evaluation progress. Based on the experience of this investigation, it is recommended that further research include an approach to data collection concurrent with the undertaking of the evaluation.

Some concern that the collection of data from stakeholders at only three points might lead to some difficulties in the clear delineation of causality guided the structure of the Stakeholder Interview Schedule II. For example, the final stakeholder interviews collected information regarding the stakeholders' commitment to the study post evaluation and their perceptions of the use of the evaluation and its information. The second Stakeholder Interview schedule was structured in such a way to overcome the problem of determining causality from data collected at the same point in time. For instance, the items focusing on stakeholder commitment were asked prior to items addressing the use of the evaluations' information. Furthermore, the fact that both Vlahov (1990) and Cummings (1997) conducted studies with similar approaches supported the choice of data collection strategy.

The choice to collect both types of information in the one interview was made with consideration of resources available to the study and the availability of the stakeholders and of their time. In terms of causality, shorter interviews conducted with stakeholders more frequently throughout the study process would have enabled clearer insight into stakeholder perceptions of the evaluation process and its information, and the influence they have on the perceptions of stakeholders.

The findings of this study have a number of implications for practice with regards to undertaking an evaluation study and these will be discussed in a later section. The findings, however, can only be interpreted appropriately when the design weaknesses are kept in mind. The primary design weaknesses for the study are the difficulties resulting from the collection of data from

three independent subject groups (evaluators, stakeholders, and representatives of the organisation responsible for program delivery), the difficulties encountered in measuring the concepts of program theory, the compromises to the values of statistical analysis due to the small sample size, and the inclusion of only three cases for investigation. Therefore, some caution needs to be taken in consideration of the findings of the study and, in turn, the conclusions drawn from these findings and the implications for evaluation practice. These design weaknesses, discussed further below, may be overcome in future research to test the findings indicated in this study.

Data management techniques, necessitated by the collection of interview data from three distinct interview subject groups, likely led to the loss of important empirical information regarding the influence of the organisation environment on the evaluation process and its outcomes. This is also the case for the *Evaluator Characteristic* and *Stakeholder Characteristics* predictor variables included in the Fully Elaborated Structural Model. The inclusion of a large subject group from which to collect the data of any future research is important for valid and reliable statistical analysis of study data. The problem of having only a small subject group is that in the assessment of statistical analysis caveats must be made regarding the possible compromise to the results, due to the calculations being based on small sample size. This is particularly true for structural equation modelling.

The intention of the current study was to investigate three independent evaluation studies with a view to comparing and contrasting the influence of different types and degrees of program theory use in the evaluation. Although an extensive search, using both formal and informal networks, was conducted to identify evaluation studies which fitted with the conditions for inclusion in the study, and a large number of studies were considered, the number of available evaluation studies to choose from was small. The result of having such a limited selection to choose from resulted in two of the studies having some overlap in terms of the evaluators responsible for the study conduct. Furthermore, the concern of having so few studies is that idiosyncrasies of any one evaluation study could sway the findings of this study. In particular, it is

suspected that this is the case with regards to the YOHFest Program evaluation.

In the analysis of the data of this study, the responses of an evaluator involved in two of the studies were only included once, so as not to allow the perceptions of this person to overly influence the findings, particularly as the subject groups representing the evaluators was very small. For this reason, another caveat has had to be taken into consideration in the interpretation of the findings. This is that, with regards to future studies, every effort should be made to maintain the independence of the evaluation studies investigated with a view to strengthening the clarity of the findings.

#### IMPLICATIONS FOR FUTURE RESEARCH

In 1977 Straton raised questions about research of the evaluation process which have current relevance for this study. Straton suggested research was needed with regards to: the identification of the kinds of questions an evaluation is to address, or the *delineation* of the evaluation focus, information collection processes adopted in undertaking an evaluation, or the *obtaining* of evaluation information, the most effective way of *providing* evaluation information to the audiences, and methods of enhancing the *utilising* of the evaluation's information. Research investigating the influence a program's theory in the undertaking of an evaluation addresses all four areas of Straton's model of the evaluation process.

In particular, this study has aimed to provide insight into the effect of program theory in the undertaking of a program evaluation on the evaluation study process, the influence of the evaluation, and the use of evaluation information. The model of this study is concerned with adding to the evaluation utilisation research. In particular, the factor of program theory use in the evaluation process has been included in the structural equation model versions.

Similarly, challenges to utilising the evaluation information were identified as having an impact. For instance, two of the evaluations investigated were of

programs developed and driven from a central location, yet implemented locally. Therefore, stakeholders located away from the central office had little opportunity to put the evaluation, and the information it yielded, to use. Their role in the delivery of the program required them to respond to the decisions made in central office. This lack of opportunity to use the information was identified as influencing their motivation to read and process the information contained in the final report of the evaluation.

Implications for further research have relevance to research and practice. It is suggested that future research in this area include studies with more cases, to strengthen the findings and gain better insight into the influence of program theory on use of evaluation information. This would also enable investigation of different sorts of evaluation studies with a greater coverage and variation in the degree of use of both implementation program theory and causative program theory, with a view to comparing and contrasting the effect of each on the evaluation study process, the influence of the evaluation, and the influence of the information yielded by the evaluation study. The issue of how to measure the extent of the two types of program theory was given a great deal of consideration in this study. It is, however, a difficult concept to measure and provides opportunities for more work to be undertaken in the development of tools to assess the impact of program theory in evaluation studies.

The Classification Matrix developed in this study represents an effort to classify the source, type, degree and level of contextualisation of program theory use in an evaluation study, as no others were found in the literature. A review of literature relevant to program theory use in undertaking evaluation studies indicated these to be the four most debated issues regarding program theory (e.g. Bickman, 1987; Chen, 1990; Mertens, 1996; Patton, 1998; Pawson and Tilley, 1997; Scheirer, 1987; Weiss, 1997a; Weiss, 1997b). Future empirical research into the influence of program theory in evaluation might usefully build on the four dimensions included in the Classification Matrix developed in this study.



Conceptually, the program theory terminology is often difficult for many stakeholders to grasp, although they may have very clear tacit knowledge of a program's theory. Therefore, the task of determining what a program's stakeholders know and understand to be a program's theory, is an area open to further research, particularly with regards to the processes of program theory elaboration and the skills required of evaluators to ensure it is done well. This study considered perceptions of evaluators regarding their personal practice, in terms of undertaking an evaluation and previous experience and training in social research and evaluation. As a result, it represents an important, but limited, attempt to determine the characteristics of practice useful in evaluators, with a view to undertaking a theory-based evaluation, particularly one including a theory elaboration exercise among stakeholders.

#### IMPLICATIONS FOR EVALUATION PRACTICE

The recognition of the impact of the evaluation study characteristics on the use of an evaluation's information is the most obvious finding of this study, based upon the empirical quantitative and qualitative evidence. It seems clear that even when the situation is such that many other characteristics are apparently optimal for an evaluation to have an impact, lengthy delays in the release of the program report may bring about stakeholder disillusionment with the evaluation. The disillusionment will likely lead to a reduction in the possible impact of the evaluation study. Furthermore, the late release of the evaluation information may result in missed opportunities for information to have an influence if stakeholders are in a position where they have limited opportunity to put the evaluation information to use.

It is likely that evaluators with more experience might identify potential pitfalls to the evaluation process in advance, though often issues arise with little or no warning. The evaluators of the YOHFest Program were very experienced and knowledgeable of the program theory elaboration process. Unforeseen circumstances led to difficulties in time management on the part of the evaluators which caused the report to be late. However, with regards to the YOHFest Program evaluation in particular, qualitative responses of

stakeholders to the second stakeholder interview indicated that although they were very displeased with the way the evaluation turned out, they felt the difficulties with the evaluation were exacerbated through infrequent communication with the evaluator at the time the final report was due for release. Consideration of the brief upon which the evaluation was contracted gives further insight into why a degree of tension developed between the evaluators of the YOHFest Program and the stakeholders. The YOHFest organisation needed the program to be formally evaluated to meet funding requirements and to inform further program development. The evaluator agreed to undertake the evaluation primarily based on a briefing given to him by a friend, who also happened to be a sponsor of the YOHFest Program. However, the stakeholders actually had plans for the evaluation study that might not have been detailed as clearly as needed in the early stages of the study. In the process of the evaluation it became evident that the expectations of the stakeholders were not realistic with regards to the funds they had available for the study. It is likely more communication early in the study would have highlighted this inconsistency and more realistic stakeholder expectations may have been negotiated.

The original fully elaborated model developed in which the *Evaluator Characteristics* latent variable was tested, did not fit the data. However, it is suspected that good identification of evaluator practice characteristics, based on agreed understandings of good evaluation practice, will enable the selection and training of evaluators with ability and skill to undertake exemplary evaluations with good influence.

When the task of elaborating stakeholders' understanding of program theory is considered, the need for a highly and multi-skilled evaluator is emphasised. If stakeholders are likely to have a difficult time understanding what is meant by program theory, it is important that those responsible for leading them through the theory elaboration process, or working to understand stakeholder perceptions of what is meant by program theory, have knowledge of the characteristics of practice most useful for undertaking the task. This study has taken some initial steps in the development of tools to assess evaluator

practice, based on their perceptions of practice. It is suggested further research is conducted to develop such tools and evaluators endeavour to develop the competencies identified.

Although, the quantitative findings of this study appear to indicate that the use of causative theory in the conduct of an evaluation can have a direct negative impact on the stakeholder's perception of the evaluation study characteristics, and indirectly on the use of the evaluation information, a consideration of the qualitative data highlights the influence of other aspects of the evaluation study which are more likely to have brought about the negative effect. In fact, the responses of stakeholders of all three evaluations indicated they were all interested in understanding the theory of the program or logic of the program and felt the use of the program theory in the conduct of the evaluation was a useful way to gain insight into the program.

Pre-Apprenticeship programs stakeholders were specifically interested in the implementation theory of the program. They contracted the evaluation specifically to gain an understanding of what the program had evolved into, as they suspected the program in operation was very different from that which was originally implemented. YOHFest Program stakeholders were interested primarily in causative theory of the program and were concerned that the evaluation should provide quantitative data regarding the effect of the causative theory of the program. They specifically required this information of the evaluation to include in their application for further funding. Furthermore, YOHFest Program stakeholders involved in the program theory elaboration process at the beginning of the evaluation, all acknowledge the positive benefits of the practice in relation to the explication of their thinking regarding the program and that of others with involvement in a program. Finally, the MCJ program is underpinned by identified use of both causative theory and implementation theory and stakeholders of the program were very keen to gain insight into the implementation and effect of the program in terms of the program theory. The finding that MCJ Program stakeholders did not use the evaluation information is likely to be due to circumstances other than program theory use in the evaluation, such as limited opportunities for use.

The *Process Use* variable of the Core Model was found to have significant influence on the outcome variables. The qualitative responses of stakeholders confirm they found involvement in the evaluation to have influenced their thinking, practice, or decisions they were responsible for making. Therefore, it seems prudent to recognise the influence of stakeholder involvement in the evaluation and the likelihood of an evaluation having influence midway through its undertaking. Furthermore, the impact of stakeholder involvement, as measured by the *Stakeholder Involvement (Post)* latent variable, appears to have had a significant direct positive effect on the extent to which stakeholders are influenced by the evaluation information.

Finally, the empirical findings of this study indicate that the extent to which the environment of the organisation responsible for the delivery of the program is characteristics of a learning environment has a negative impact on the use of the information. In fact, this finding is likely due to the small number of cases upon which this investigation is based. The YOHFest *Organisational Learning Environment* scores were highest, while the other two representative organisations scored roughly the same with regards to the organisational environment. It is very likely that the issues with the YOHFest evaluation process, identified previously, detracted from greater use of the evaluation information, not the extent to which the YOHFest organisational environment is characteristic of a learning environment. In fact, in the initial interviews with stakeholders of the YOHFest Program, their responses indicated they were very keen to access new information and very proactive in unearthing and following through on opportunities to positively develop their program with a view to encouraging the adoption of healthy lifestyle attitudes and practices among youth. They had already identified clear avenues for the evaluation's information to have influence. With this in mind, it is suggested that an awareness of the nature of the environment in which the evaluation's information is likely to have influence is acknowledged in the undertaking of an evaluation.

## CONCLUSION

The present study contains a number of implications for evaluation practice. However, care should be taken in the direct application of the implications, as this study has some weaknesses which have lessened the strength of the findings in terms of their validity. These weaknesses include the small sample size upon which the study results are based, the difficulties which resulted from the collection of data from three independent subject groups (evaluators, stakeholders, and representatives of the organisation responsible for program delivery), and the investigation of only three program evaluations enabling greater opportunity for each case to influence the overall findings of the study than would have been the case had the study included a greater number.

The results of this study indicate the use of program theory in the undertaking of an evaluation study has no direct impact on the use of the evaluation's information. However, it does seem to interact with some other predictor variables, such as *Evaluation Study Characteristics*. With regards to the cases of this study, there are a number of contextual factors to be considered in the interpretations of the findings which may have affected the impact of program theory on the use of the evaluation information. Qualitative information gathered in the study gives insight regarding the vulnerability of a study to contextual factors, often outside the control of evaluators. Alternatively, program theory use in the evaluation was found to have a significant indirect impact on the use of evaluation information through the stakeholder's perceptions of the characteristics of the evaluation study.

This study has not been able to confirm claims by those such as Bickman (1987b), Chen (2004), and Weiss (1998) that a more valuable and useful study will result if an evaluation is based on program theory. While Chen states that the "...benefits that program theory provides to evaluation are well documented in the evaluation literature" (Chen, 2004, p 15) he refers to the discussion papers of Bickman (1987) and Weiss (1997a; 1998), not empirical studies of the influence of program theory-based evaluations. In fact, Rogers et al. (2000) suggest a lack of "real-world test and applications" of theory-

based evaluations indicates evaluation practitioners are working from professional intuition, rather than empirical evidence. This study highlights the need for evaluation practice to be subjected to empirical tests to validate the practice. Good quality empirical research will enable evaluation practitioners to have greater confidence that they are maximising the likelihood the information provided by their evaluations will be used effectively.

## REFERENCE LIST

- Ajzen, I. and Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Englewood Cliffs, NJ: Prentice Hall.
- Alkin, M. (1990). *Debates on Evaluation*. Newbury Park, California: Sage.
- Alkin, M. (2004). Context-Adapted Utilisation: A Personal Journey. In A. Alkin (Ed.), *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage.
- Alkin, M. and Associates (1985). *A Guide for Evaluation Decision-Makers*. Beverly Hills, California: Sage.
- Alkin, M., Daillak, R. and White, P. (1979). *Using Evaluations: Does Evaluation Make a Difference?*, Beverly Hills, California: Sage.
- Alkin, M. and Taut, S. (2003). Unbundling Evaluation Use. *Studies in Educational Evaluation*, 29 (1), 1-12.
- Anderson, J.C. and Gerbing, D. (1984). The effects of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155-173.
- Arbuckle, J. 1995-2005. AMOS<sup>tm</sup> 6.0 User's Guide. AMOS Development Corporation: Spring House, PA.
- Argyris, C. (1999). *On Organisational Learning* (2nd ed). Oxford: Blackwell.
- Argyris, C., Putnam, R. and Smith, D. (1985). *Action Science*. San Francisco, CA: Jossey-Bass.
- Argyris, C. and Schön, D. (1978) *Organisational Learning*. Massachusetts: Addison-Wesley.
- Atweh, B., Kemmis, S. and Weeks, P. (Eds.). (1998). *Action Research in Practice: Partnerships for Social Justice in Action*. London: Routledge.
- Babbie, E. (1998). *The Practice of Social Research*. 8th ed. Belmont, California: Wadsworth Publishing Company.
- Bandura, A. (1977). *Social Learning Theory*. New Jersey: Prentice-Hall.
- Baron, R. and Kennedy, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51 (6), 1173- 1182.
- Bartlett, C. and Ghoshal, S. (1995). Rebuilding behavioural context: Turn process reengineering into people rejuvenation. *Sloan Management Review*, 11 - 23.
- Bennett, T. (1996). What's new in evaluation research? A note on the Pawson and Tilley article. *British Journal of Criminology*, 36, 567-573.
- Bentler, P. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-246.

- Bentler, P. (1995). *EQS: Structural equations program manual*. Encino, CA: Multivariate Software, Inc.
- Bentler, P. and Bonett, D. (1980). Significance test and goodness-of-fit in the analysis of co-variance structures. *Psychological Bulletin*, 88, 588-606.
- Bentler, P. and Chou, C. (1987). Practical issues in structural modelling. *Sociological Methods and Research*, 16, 78-117.
- Bickman, L. (1987). The importance of program theory. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 33, 5-18. San Francisco: Jossey-Bass.
- Bickman, L. (Ed.). (1987b). *Using Program Theory in Evaluation*. San Francisco: Jossey Bass.
- Bickman, L. (1989). Barriers to the use of program theory. *Evaluation and Program Planning*, 12, 387-390.
- Bickman, L. (1996). The application of program theory to the evaluation of a managed mental health care system. *Evaluation and Program Planning*, 19 (2), 111-119.
- Bickman, L. and Petersen, K. (1990). Using program theory to describe and measure program quality. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 47, 61-72.
- Biglan, A., Ary, D., Yudelson, H., Duncan, T., Hood, D., James, L., Koehn, V., Wright, Z., Black, C., Levings, D., Smith, S., and Gaiser, E. (1996). Experimental Evaluation of a modular approach to mobilizing antitobacco influences of peers and parents. *American Journal of Community Psychology*, 24 (3), 311-339.
- Bloom, B. Hastings, T. and Madaus, G. (1970). *Handbook on Formative and Summative Evaluation*. New York: McGraw-Hill.
- Bollen, K. (1986). Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika*, 51, 375-377.
- Bollen, K. (1989). A new incremental fit index for general equation models. *Sociological Methods and Research*, 17, 303-316.
- Bollen, K. and Long, J. (Eds.). (1993). *Testing Structural Equation Models*. Newbury Park, CA: Sage.
- Braskamp, L. and Brown, R. (1980). *Utilisation of Evaluation Information*. San Francisco, California: Jossey-Bass.
- Braskamp, L., Brown, R. and Newman, D. (1982). Studying evaluation through simulations. *Evaluation Review*, 6, 114-126.
- Brekke, J. (1987). The model-guided method for monitoring program implementation. *Evaluation Review*, 11 (3), 281-299.
- Brett, B., Hill-Mead, L. and Wu, S. (2000). Perspectives on evaluation use and demand by users: The case of city year. In V. Caracelli and H. Preskill (Eds.), *The Expanding Scope of Evaluation Use. New Directions for Program Evaluation*, 88, 71-85.



- Browne and Cudeck (1989). single sample cross-validation indices for covariance structures. *Multivariate Behavioural Research*, 24, 445-455.
- Bryant, D. and Bickman, L. (1996). Methodology for evaluating mental health case management. *Evaluation and Program Planning*, 19 (2), 121-129.
- Byrne, B. (1989). *A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models*. New York: Springer-Verlag.
- Byrne, B. (1995). One application of structural equation modelling from two perspectives: Exploring the EQS and LISREL strategies, In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (138-157). Thousand Oaks: Sage.
- Byrne, B. (2001). *Structural Equation Modelling with AMOS: Basic Concepts, Applications, and Programming*. New Jersey: Lawrence Erlbaum.
- Cabatoff, K. (1996). Getting on and off the policy agenda: A dualistic theory of program evaluation utilisation. *The Canadian Journal of Program Evaluation*, 11(2), 35-60.
- Calsyn, R., Roades, L., and Klinkenberg, W. (1998). Using theory to design needs assessment studies of the elderly. *Evaluation and Program Planning*, 21, 277-286.
- Campbell, D. (1969). Reforms as experiments. *American Psychologist*, 24, 409-429.
- Campbell, D. (1971). Methods for the experimenting society. Paper presented at the meeting of the American Psychological Association, Washington, D.C.
- Carmine, E. and McIver, J. (1981). Analysing models with unobserved variables. In G.W. Bohrnstedt and E.F. Borgatta (Eds.), *Social Measurement: Current Issues*. Beverly Hills: Sage Publications.
- Caro, F.G. (1971). Issues in the evaluation of social programs. *Review of National Research*, 41, 87-114.
- Chen, H. (1989a). Guest editor's note: The theory-driven perspective. *Evaluation and Program Planning*, 12, 297.
- Chen, H. (1989b). The conceptual framework of the theory-driven perspective. *Evaluation and Program Planning*, 12, 391-396.
- Chen, H. (1990a). Issues in constructing program theory. In L Bickman, (Ed.), *New Directions for Program Evaluation*, 47, 7-18.
- Chen, H. (1990b). *Theory-Driven Evaluations*. Newbury Park, CA: Sage.
- Chen, H. (1994a). Current trends and future directions in program evaluation. *Evaluation Practice*, 15, 229-238.
- Chen, H. (1994b). Epilogue. *Evaluation Practice*, 15, 93-93.
- Chen, H. (1994c). A panel of theory-driven evaluation and evaluation theories. *Evaluation Practice*, 15, 73-74.
- Chen, H. (1994d). Theory-driven evaluations: Need, difficulties, and options. *Evaluation Practice*, 15, 79-82.

- Chen, H. (2003). Theory-driven approach for facilitation of planning health promotion or other programs. *The Canadian Journal of Program Evaluation*, 18 (2), 91-113.
- Chen, H. (2004). *Practical Program Evaluation, Assessing and Improving Planning, Implementation, and Effectiveness*. London: Sage.
- Chen, H. and Rossi, P. (1981). The multi-goal, theory-driven approach to evaluation: A model lining basic and applied social science. In H. Freeman and M. Solomon (Eds.), *Evaluation Studies Review Annual*, 6, 38-54.
- Chen, H. and Rossi, P. (1989). Issues in the theory-driven perspective. *Evaluation and Program Planning*, 12, 299-306.
- Chen, H. and Rossi, P. (1992). Introduction: Integrating theory into evaluation practice. In H. Chen and P. Rossi (Ed.), *Using Theory to Improve Program and Policy Evaluation*. Westport, CT: Greenwood.
- Chen, H., Wang, J., and Lin, L. (1997). Evaluating the process and outcome of a garbage reduction program in Taiwan. *Evaluation Review*, 21(1), 27-42.
- Chou, C., and Bentler, P. (1995). Estimates and tests in structural equation modelling, In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (37-55). Thousand Oaks: Sage.
- Cohen, D. and Rice, J. (1995). A parent-targeted intervention for adolescent substance use prevention. *Evaluation Review*, 19 (2), 159-180.
- Cohen, J. (1965). Some statistical issues in psychological research. In B..B. Wolman (Ed.), *Handbook of Clinical Psychology* (95-121). New York: McGraw-Hill.
- Collins, C., Kohler, C., DiClemente, and Wang, M.Q. (1999). Evaluation of the exposure effects of a theory-based street outreach HIV intervention on African-American drug users. *Evaluation and Program Planning*, 22, 279-293.
- Conley-Tyler, M. (2005). A fundamental choice: Internal and external evaluation. *Evaluation Journal of Australasia*, 4 (new series), 1 and 2, 3-11.
- Conrad, K. and Buelow, J. (1990). Developing and testing program classification and function theories. In L. Bickman (Ed.), *New Directions for Program Evaluation*, no. 47, 73-91.
- Conrad, K. and Miller, T. (1987). From program theory to tests of program theory. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 33, 19-42.
- Constantine, N. and Curry, K. (1998, November). Collaborative development of a theory-based student assessment for a violence prevention program evaluation. Paper presented at the American Evaluation Association Annual Conference, Chicago, Illinois.
- Cook, T. and Campbell, D. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Beverly Hill, CA: Sage.
- Cook, T. and Reichardt, C. (Eds.). (1979). *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills, CA: Sage.

- Cook, T., Leviton, L., and Shadish, W. (1985). Program evaluation. In G. Lindzey and E. Aronson (Eds.), *Handbook of Social Psychology*. (3<sup>rd</sup> ed., 699-777). New York: Random House.
- Cousins, B. and Earl, L. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14 (4), 397-418.
- Cousins, B. and Earl, L. (Eds.). (1995). *Participatory Evaluation in Education: Studies in Evaluation Use and Organisational Learning*. London: Falmer.
- Cousins, J. (1996). Understanding organisational learning for educational leadership and school reform. In K. Leithwood, J. Chapmans, D. Corson, P. Hallinger, and A. Hart (Eds), *International Handbook of Educational Leadership and Administration* (589-652). Boston: Kluwer Academic Publishers.
- Cousins, J. and Leithwood, K. (1986). Current empirical research on evaluation utilisation. *Review of Educational Research*, 56, 331-364.
- Cousins, J. and Leithwood, K. (1993). Enhancing knowledge utilisation as a strategy for school improvement. *Knowledge: Creation, Diffusion and Utilisation*, 14, 305-333.
- Cousins, J. and Walker, C. (2000). Predictors of Educators' valuing of systematic inquiry in schools. *The Canadian Journal of Program Evaluation*, Special Issue, 25-52.
- Cronbach, L. (1963). *Course improvement through evaluation*. Teacher's College Record, 64, 672-683.
- Cronbach, L., Ambron, S., Dornbusch, S., Hess, R., Hornik, R., Phillips, D., Walker, D. and Weiner, S. (1980). *Toward reform of program evaluation*. San Francisco, CA: Jossey-Bass.
- Cronbach, L. and Snow, R. (1977). *Aptitudes and Instructional Methods: A Handbook for research on Interactions*. New York: Irvington.
- Cummings, O., Nowakowski, J., Schwandt, T., Eichelberger, R., Kleist, K. Larson, C. and Knott, T. (1988). Business perspectives on internal/external evaluation. In J. McLaughlin, L. Weber, R. Covert and R. Ingle (Eds.), *Evaluation Utilisation*, 39, 59-74.
- Cummings, R. (1997). *The Influence of Stakeholder Involvement in Evaluation Studies on the Use of Evaluation Information: A Longitudinal Study*. Unpublished Doctor of Philosophy dissertation, Murdoch University, Western Australia.
- Cummings, R., Stephenson, K., and Hale, L. (2001). Using program theory in an education setting. *Evaluation Journal of Australasia*, 1 (1) (new series), 29-39.
- Cummings, R., and Coldrick, B. (2004, May). *Evaluation Study of the CIP2 Professional Learning Program: Final Report*. Perth, Western Australia: Estill and Associates/ Murdoch University.

- Cummings, R., Aquilina, H., and Coldrick, B. (2005, June). *CIP2 Professional Learning Program Evaluation: Final Report*. Perth, Western Australia: Estill and Associates/ Murdoch University.
- Cummings, R., Kapiteyn, A. (2005). *Evaluation of YohFest 2004: Final Report 2005*. Perth, Western Australia: Estill and Associates/ Murdoch University.
- Dagenais, F. (1978). Theory-based evaluation of an experimental program. *Educational Research Quarterly*, 3, 74-84.
- Davis, H. and Salasin, S. (1975). The utilisation of evaluation. In E.L. Struening and M. Guttentag (Eds.), *Handbook of Evaluation Research* (621-666), Beverly Hills, California: Sage.
- Dawson, J., and D'Amico, J. (1985). Involving program staff in evaluation studies: A strategy for increasing information use and enriching the data base. *Evaluation Review*, 9, 173-188.
- de Mello Vianna, F. (1980). *Roget's II The New Thesaurus*. Boston, MA: Houghton Mifflin Company.
- de Vaus, D. (1985). *Studies in Society: Surveys in Social Research* ( 2<sup>nd</sup> ed.), Sydney: Allen and Unwin.
- Department of Education and Training. (2004). *Request for quotation project brief: An evaluation of Pre-Apprenticeship Training Programs in Western Australia*. Perth, Australia.
- Department of Education and Training. (2004, June). *Statement on Professional Learning (Draft)*. Curriculum Directorate, Department of Education and Training, Perth Western Australia.
- Department of Training and Employment. (2003). *Enrolment in TAFE Pre-apprentice Courses (What has changed from 2001 to 2002?)*. Perth, Australia.
- Dickey, B. (1980). Utilisation of evaluations of small-scale innovative educational projects. *Educational Evaluation and Policy Analysis*, 2, 65-77.
- Dickman, F. (1981). Work activities, settings, methodologies and perceptions: Correlates of evaluative research utilisation. *Knowledge: Creation, Diffusion, Utilisation*, 2, 375-387.
- Dilworth, R. (1995). The DNA of the learning organisation. In S. Chawla and J. Renesch (Eds.), *Learning Organisations: Developing Cultures for Tomorrow's Work Place*, Oregon: Productivity Press.
- English, B. (2002). Competencies for evaluation practitioners: Where to from here? *Evaluation Journal of Australasia*, 2(2).
- Estill and Associates/ Murdoch University. (2005, June). *Evaluation Study of the CIP2 Professional Learning Program: Final Report*. Perth, Western Australia.
- Fetterman, D. (1988). *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution*. Westport, CT: Praeger.

- Fetterman, D. (1992). Theory in evaluation: We think, therefore we theorize (An ethnographer's perspective). In H. Chen and P. Rossi (Ed.), *Using Theory to Improve Program and Policy Evaluation*. Westport, CT: Greenwood.
- Fetterman, D. (1994). Empowerment evaluation. *Evaluation Practice*, 15, 1-15.
- Fetterman, D. (2004). Branching Out or Standing on a Limb: Looking to our Roots for Insight. In M. Alkin (Ed.), *Evaluation Roots: Tracing Theorists' Views and Influences* (304). Thousand Oaks, CA: Sage.
- Finney, J. and Moos, R. (1992). Four types of theory that can guide treatment evaluation. In H. Chen and P. Rossi (Ed.), *Using Theory to Improve Program and Policy Evaluation*. Westport, CT: Greenwood.
- Fitz-Gibbon, C. and Morris, L. (1972). Theory-based evaluation. *Evaluation Comment*, 5 (1), 1-4.
- Fitz-Gibbon, C. and Morris, L. (1996). Theory-Based evaluation. *Evaluation Practice*, 17(2), 177-184.
- Funnell, S. (1997). Program Logic: An adaptable tool for designing and evaluating programs. *Evaluation News and Comment*, 6 (1), 5-17.
- Garvin, D. (1993). Building a learning organisation. *Harvard Business Review*, July-August, 78-91.
- Glaser, B. and Strauss, A. (1967). *The discovery of grounded theory: strategies for qualitative research*. Chicago : Aldine.
- Glaser, B. G. (2000). *The discovery of grounded theory*. Mill Valley, CA: Sociology Press.
- Goldberger, A. (1973). Structural equation models: An overview. In A. S. Goldberger and O.D. Duncan (Eds.), *Structural Equation Models in the Social Sciences* (1-18). New York: Seminar Press.
- Goodman, R., Wandersman, A., Chinman, M., Imm, P., and Morrissey, E. (1996). An ecological assessment of community-based interventions for prevention and health promotion: Approaches to measuring community coalitions. *American Journal of Community Psychology*, 24 (1), 33-61.
- Grant, C. (1990). The promise and perils of praxis. *Cross Currents*, 40, 64-87.
- Greene, J. (1987). Stakeholder participation in evaluation design: Is it worth the effort? *Evaluation and Program Planning*, 10 (4), 379-394.
- Greene, J. (1988). Stakeholder participation and utilisation in program evaluation. *Evaluation Review*, 12, 91-116.
- Greene, J. (2002). Mixed-method evaluation: a way of democratically engaging with difference. *The Evaluation Journal of Australiasia*, 2 (2), 23-28.
- Greene, J. and Kvidahl, R. (1990, April). *Research Methods Courses Post-Bachelor's Education: Effects on Teachers' Research Use and Opinions*. Paper presented at American Educational Research Association, Boston.

- Guba, E. (1969). The failure of educational evaluation. *Educational Technology*, 9, 29-38.
- Guba, E. and Lincoln, Y. (1981). *Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches*. San Francisco, CA: Jossey-Bass.
- Guskey, T. (1988). Teacher efficacy, self-concept, and attitudes towards the implementation of instructional innovation. *Teaching and Teacher Education*, 4(1), 63-69.
- Hamilton, D., Jenkins, D., King, C., MacDonald, B., and Parlett M. (1977). *Beyond the Numbers Game*. Great Britain: Macmillan Education.
- Hammond, J. (1982). *The personal knowledge factor and the utilisation of evaluation*. Unpublished Master of Education dissertation, Murdoch University, Western Australia.
- Harre, R. (1972). *The Philosophies of Science*. Oxford: Oxford University Press.
- Hayduk, L. (1987). *Structural equation modelling with LISREL: Essentials and advances*. Baltimore: The John Hopkins University Press.
- Henry, G.T. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation*, 24(3), 293-314.
- Henry, G.T. (2004). The Mechanisms and Outcomes of Evaluation Influence. *Evaluation*, 10(1), 35-57.
- Henry, G.T. and Mark, M. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation*, 24, No. 3, 293-314.
- Herman, S. (1997). Exploring the link between service quality and outcomes: Parents' assessments of family support programs. *Evaluation Review*, 21 (3), 388-404.
- Hinkle, D., Wiersma, W., and Jurs, S. (1988). *Applied Statistics for the Behavioural Sciences* ( 2<sup>nd</sup> ed.). Boston: Houghton.
- Holmes-Smith, P. (2001). *Introduction to Structural Equation Modelling Using AMOS: Course Outline and Resources*. ACSPRI 2001 (Summer Program); Canberra.
- Holmes-Smith, P. (July 24, 2007). Personal communication.
- House, E. and Howe, K. (1998). Advocacy in Evaluation. *American Journal of Evaluation*, 19, 233-236.
- House, E. (1980). *Evaluating with Validity*. Beverly Hills: Sage Publications.
- Hoyle, R. (1995). The structural equation modelling approach, In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (1-15). Thousand Oaks: Sage.

- Hoyle, R. and Panter, A. (1995). Writing about structural equation models. In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (158-176). Thousand Oaks: Sage.
- Hu, L. and Bentler, P. (1995). Evaluating model fit, In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (76-99). Thousand Oaks: Sage.
- Hudson-Mabbs, S. (1993). *Influences on the use of evaluation information: Three case studies*. Unpublished Master of Education dissertation, Murdoch University.
- Hudson-Mabbs, S. and Straton, R. (1999). Factors affecting information utilisation and implications for evaluation practice. *Evaluation Journal of Australasia*, 11(2), 64-79.
- Huebner, T. (2000). Theory-based evaluations: Gaining a shared understanding between school staff and evaluators. In P. Rogers, T. Hasci, A. Petrosino, T. Huebner (Eds.), *Program Theory in Evaluation: Challenges and Opportunities. New Directions for Evaluation*, 87, 79-90. San Francisco, Jossey-Bass.
- Hughes, J. (1990). *The Philosophy of Social Research*. Essex, England: Longman.
- Hull, J. Tedlie, J. and Lehn, D. (1995). Modelling the relation of personality variables to symptom complaints; The unique role of negative affectivity, In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (217-235). Thousand Oaks: Sage.
- Information Technology Services at the University of Texas at Austin. (2002). Structural Equation Modelling Using AMOS: An Introduction. Last updated December 15, 2002. Retrieved January 10, 2007 from <http://www.utexas.edu/its/rc/tutorials/stat/amos/>.
- James, L., Mulaik, S., and Brett, J. (1982). *Causal Analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.
- Jenlink, P. (1994). Using evaluation to understand the learning architecture of an organisation. *Evaluation and Program Planning*, 17(3), 315-325.
- Johnson, D., and Johnson, R. (1975). *Learning Together and Alone: Cooperation, Competition, and Individualization*. Englewood Cliffs, N.J.; Sydney: Prentice-Hall
- Johnson, R. (1998). Towards a theoretical model of evaluation utilisation. *Evaluation and Program Planning*, 21, 93-110.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for Evaluations of Educational Programs, Projects, and Materials*. New York: McGraw Hill.
- Joint Committee on Standards for Educational Evaluation. (1994). *The Program Evaluation Standards* (2<sup>nd</sup> ed.), Thousand Oaks, CA: Sage.
- Julian, D. (1997). The utilisation of the logic model as a system level planning and evaluation device. *Evaluation and Program Planning*, 20, 251-257.

- Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- Jöreskog, K. and Sörbom, D. (1984). *LISREL VI user's guide* (3<sup>rd</sup> ed.). Mooresville, IN: Scientific Software.
- Jöreskog, K. and Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS.
- Kemmis, S. and McTaggart, R. (2000). Participatory Action Research. In N. Denzin and Y. Lincoln (Eds.), *The Handbook of Qualitative Research* (2nd ed.). Beverly Hills: Sage.
- Kerlinger, F. (1979). *Behavioural Research: A Conceptual Approach*. New York: Holt, Rinehart and Winston.
- Kiesler, C.A. (1971). *The psychology of commitment*. New York: Academic Press.
- Kim, D. H. (1993). The link between individual and organisational learning. *Sloan Management Review*. Fall, 37-50.
- Kirby, D., Korpi, M., Adivi, C., and Weissman, J. (1997). An impact evaluation of Project SNAP: An AIDS and pregnancy prevention middle school program. *AIDS Education and Prevention*, 9, 44-61.
- Kirkhart, K. (2000). Reconceptualising Evaluation Use: An Integrated Theory of Influence. In V. Caracelli and H. Preskill (Eds.), *The Expanding Scope of Evaluation Use*. *New Directions for Evaluation*, 88, 5-23. San Francisco, Jossey-Bass.
- Lake, R. (2005). Fly high, go deep, add value: Characteristics of 'black belt' evaluators. *Evaluation Journal of Australasia*, 5 (new series), 2, 27-31.
- Lather, P. (1987). Research as praxis. In W. Shadish and C. Reichardt (Eds.), *Evaluation Studies Review Annual*, 12, 437-457. Beverly Hills: Sage.
- Lave, C. and March, J. (1993). *An Introduction to Models in the Social Sciences* (2nd ed.). Lanham, MD: University Press of America.
- Leithwood, K. and Louis, K. (1999). *Communities of Learning and Learning Schools: New Directions for School Reform*. Amsterdam: Swets and Zeitlinger.
- Leslie, G., Larson, R., and Gorman, B. (1973). *Order and Change: Introductory Sociology*. New York: Oxford University Press.
- Levin, B. (1987). The uses of research: a case study in research and policy. *The Canadian Journal of Program Evaluation*, 2(1), 43-55.
- Leviton, L. (1994). Program theory and evaluation theory in community-based programs. *Evaluation Practice*, 15, 89-92.
- Leviton, L. C. and Hughes, E. (1981). Research on the utilisation of evaluations: A review and synthesis. *Evaluation Review*, 5(4), 525-548.
- Lipsey, M. (1997). What can you build with a thousand bricks: Musings on the cumulation of knowledge in program evaluation. In D. Rog and D. Fournier (Eds.), *New Directions for Program Evaluation*, 27, 7-28.



- Lipsey, M. and Pollard, J. (1989). Driving toward theory in program evaluation: More models to chose from. *Evaluation and Program Planning*, 12, 317-328.
- Lipsey, M., Crosse, S., Dunkle, J., Pollard, J. and Stobart, G. (1985). Evaluation: The state of the art and the sorry state of the science. In D. Cordray (Ed.), *New Directions for Program Evaluation*, 27, 7-28.
- Louis, K. and Simsek, H. (1991, October). *Paradigm Shifts and Organisational Learning: Some Theoretical Lessons for Restructuring Schools*. Paper presented at the annual meeting of the University Council for Education Administration, Baltimore, MD.
- Luthans, F. (1998). *Organisational Behaviour* (8th ed.), McGraw Hill: Massachusetts.
- Madaus, G. and Stufflebeam, D. (1989). *Educational Evaluation: Classic works of Ralph W. Tyler*. Boston: Kluwer Academic Publishers.
- Madaus, G., Scriven, M., and Stufflebeam, D. (1983). *Evaluation models: viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff, 1983.
- Madison, A. (1992). Primary inclusions of culturally diverse minority program participants in the evaluation process. In A. Madison (Ed.), *Minority Issues in Program Evaluation*. (35-44). San Francisco: Jossey-Bass Publishers.
- Mark, M. (1990). From program theory to tests of program theory. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 47, 37-51.
- Mark, M.M. and Bickman, L. and Petersen, K. (1990). Using program theory to describe and measure program quality. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 47, 61-72.
- Mark, M., Hofmann, D. and Reichardt, C. (1992). Testing theories in theory-driven evaluation: (Tests of) Moderation in all things. In H. Chen and P. Rossi (Eds.), *Using Theory to Improve Program and Policy Evaluation*. Westport, CT: Greenwood.
- Marsh, H., Balla, J. and McDonald R. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-411.
- Marsh, H. and Hocevar. 1985. Applications of confirmatory factor analysis to the study of self-concept: First and higher-order factor models and their invariance across groups. *Psychological Bulletin*, 97: 562-582.
- Martinson, R. (1974). What works? Questions and answers about prison reform. *Public Interest*, 35: 22-45.
- Mathison, S. (1994). Rethinking the evaluator role: Partnerships between organisations and evaluators. *Evaluation and Program Planning*, 17(3), 299-304.
- McGraw, S., Sellers, D., Stone, E., Bebachuk, J., Edmundson, E., Johnson, C., Bachman, K. and Luepker, R. (1996). Using process data to explain outcomes: An illustration for the child and adolescent trial for cardiovascular health (CATCH). *Evaluation Review*, 20, 291-312.

- McGuire, W.J. (1969). The nature of attitudes and attitude change. In G. Lindzey and E. Aronson (Eds.), *The handbook of social psychology*. Reading, Massachusetts: Addison-Wesley.
- McLaughlin, J. and Jordan, G. (1999). Logic models: a tool for telling your program's performance story. *Evaluation and Program Planning*, 22(2), 65-72.
- McLaughlin, J. and Jordan, G. (2004). Using logic models. In J. Wholey, H. Hatry and K. Newcomer (Eds.) *Handbook of Practical Program Evaluation*. Wiley, San Francisco, California, 7-32.
- McTaggart, R., Caulley, D., and Kemmis, S. (1991). Evaluation traditions in Australia: Distillation of the old, wellspring of the new. *Evaluation and Program Planning*, 14, 123-130.
- MacCallum, R. (1995). Model specification: Procedures, strategies, and related issues. In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (16-36). Thousand Oaks: Sage.
- Meltsner, A. (1976). *Policy Analysts in the Bureaucracy*. Berkeley: University of California Press.
- Mertens, D. (1996). Breaking through the silence about sexual abuse of deaf youth. *American Annals of the Deaf*, 141, (5), 352-358.
- Milligan, S., Nario-Redmond, M., and Coulton, C. (1997). *The 1995-1996 Cleveland Community-Building Initiative Baseline Progress Report: Village Council Formation, Asset Appraisal, Agenda Formation, and Action Project*. Cleveland, Ohio: Center on Urban Poverty and Social Change, Mandel School of Applied Social Sciences, Case Western Reserve University.
- Milne, C. (1993). "Outcomes Hierarchies and Program Logic as Conceptual Tools: Five Case Studies" Paper presented at the international conference of the Australasian Evaluation Society, Brisbane.
- Mulaik, S., James, L., Van Alstine, J., Bennett, N., Lind, S. and Stilwell, C. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105, 430-445.
- National Center for Vocational and Educational Research (NCVER). (2003). *Pathways to apprenticeships*. Tom Drumbell. Australia.
- Newman, D. Brown, R., Rivers, L. and Glock, R. (1983). School board's and administrators use of evaluation information: Influencing factors. *Evaluation Review*, 7(1), 110-125.
- Nutter and Nutter Associates Ltd. (1995). *Northeast In-Home Care Evaluation Project*. Paper prepared for Alberta Family and Social Services, Northeast Region.
- Owen, J. (1991). An evaluation approach to training using the notion of form: An Australian example. *Evaluation Practice*, 12 (2), 131-137.
- Owen, J. (2006). *Program Evaluation: Forms and Approaches* (3<sup>rd</sup> ed). Sydney: Allen and Unwin Pty Ltd.

- Owen, J. and Lambert, R. (1995). Roles for evaluation in learning organisations. *Evaluation*, 1(2), 237-250.
- Palumbo, D and Petersen, R. (1994). Evaluating criminal justice programs: Using policy as well as program theory. *Evaluation and Program Planning*, 17, 159-164.
- Parlett, D. and Hamilton, D. (1977). Evaluation as illumination: A new approach to the study of innovatory programmes. In D. Hamilton, D. Jenkins, C. King, B. MacDonald and M. Parlett (Eds.), *Beyond the Numbers Game*. Great Britain: Macmillan Education (6-22).
- Patrick, D., Grembowski, D., Durham, M., Beresford, S. (1999). Cost and outcomes of Medicare reimbursement for HMO preventative services. *Health Care Financing Review*, 20 (4), 25-43.
- Patton, M. (1978). *Utilisation-focused evaluation*. Beverly Hills, California: Sage.
- Patton, M. (1989). A context and boundaries for a theory-driven approach to validity. *Evaluation and Program Planning*, 12, 375-377.
- Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods* (2<sup>nd</sup> ed.). Newbury Park: Sage.
- Patton, M. (1997). *Utilisation-focused Evaluation: The New Century Text*. Thousand Oaks, CA: Sage.
- Patton, M. (1998). *Discovering Process Use*. *Evaluation*, 4(2), 225-233.
- Patton, M. (1997). *Utilisation-focused evaluation: The new century text* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Patton, M.Q. (2002). *Qualitative Research and Evaluation Methods* (3<sup>d</sup> ed.). Thousand Oaks: Sage.
- Patton, M. (2004). The roots of utilisation-focused evaluation. In M.C. Alkin (Ed.), *Evaluation Roots: Tracing Theorists' Views and Influences* (276-292). Thousand Oaks, CA: Sage.
- Patton, M., Grimes, P., Guthrie, K., Brennan, N., French, B., and Blyth, D. (1977). In search of impact: An analysis of the utilisation of federal health evaluation research. In C.H. Weiss (Ed.), *Using Social Research in Public Policy Making* (144-136). Lexington Massachusetts: D.C. Heath.
- Pawson, R. and Tilley, N. (1994). What works in evaluation research? *British Journal of Criminology*, 34, 291-306.
- Pawson, R. and Tilley, N. (1996). What's crucial in evaluation research: A reply to Bennett. *British Journal of Criminology*, 36, 574-578.
- Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage.
- Pentz, M., Trebow, E., Hansen, W., MacKinnon, D., Dwyer, J., Johnson, C., Flay, B., Daniels, S., and Cormack, C. (1990). Effects of program implementation on adolescent drug use behaviour: The Midwestern Prevention Project (MPP). *Evaluation Review*, 14 (3), 264-289.

- Petrosino, A. (2000). Answering the *why* question in evaluation: The causal model approach. *The Canadian Journal of Program Evaluation*, 15 (1), 1-24.
- Peyrot, M., Yen, S., and Baldassano, C. (1994). Short-term substance abuse prevention in jail: A cognitive behavioural approach. *Journal of Drug Education*, 24 (1), 33-47.
- Popham, W. J. (1973). The uses of instructional objectives: a personal perspective. Belmont, California: Fearon Publishers.
- Popham, W. J. (1974). Evaluation in Education: current applications. Berkeley, California: McCutchan.
- Preskill, H. and Caracelli, V. (1997). Current and developing conceptions of use: Evaluation use of TIG survey results. *Evaluation Practice*, 18 (3), 209-225.
- Preskill, H. and Torres, R. (1999a). *Evaluative Inquiry for Learning in Organisations*. Thousand Oaks: Sage.
- Preskill, H. and Torres, R. (1999b, November). Assessing an Organisation's Readiness for Learning from Evaluative Inquiry. Paper presented at the annual meeting of the American Evaluation Association, Orlando, November.
- Preskill, H. (1994). Evaluation's role in enhancing organisational learning: A model for practice. *Evaluation and Program Planning*, 17(3), 291-297.
- Preskill, H. and Torres, R. (2000a). "The learning dimension of evaluation use." In R. Torres and H. Preskill (Eds.), *The Expanding Scope of Evaluation Use*, New Directions for Evaluation, 88. San Francisco: Jossey-Bass.
- Preskill, H. and Torres, R. (2000b). The Readiness for Organisational Learning and Evaluation Instrument (ROLE). Version 4.
- Reynolds, A. (1998). Confirmatory program evaluation: A method for strengthening causal inference. *American Journal of Evaluation*, 19 (2), 203-221.
- Reynolds, A. and Walberg, H. (1990). Program theory in evaluation. In H. Walberg and G. Haertel (Eds.), *The International Encyclopedia of Educational Evaluation* (21-26). Oxford: Pergamon Press
- Rich, R. (1977). Use of social science information by bureaucrats: Knowledge for action vs. Knowledge for understanding. In C. H. Wiess (Ed.), *Uses of Social Research in Public Policy*. Lexington, Massachusetts: D.C. Heath.
- Rich, R. (1979). Problem solving and evaluation research: Unemployment insurance policy. In R. Rich (Ed.) *Translating Evaluation into Policy*. Beverly Hills: Sage.
- Rich, R. (1991). Knowledge creation, diffusion, and utilisation: Perspectives of the founding editor of Knowledge. *Knowledge: Creation, Diffusion, Utilisation*, 12, 319-337.
- Riggin, L. (1990). Linking program theory and social science theory. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 47, 109-120.

- Rippey, R. (1990). Transactional evaluation. In H. Walberg and G. Haertel (Eds.) *The International Encyclopedia of Educational Evaluation*. Oxford: Pergamon Press, 73-74.
- Rodriguez, E, and Mead, J. (1997). Evaluating a community oriented primary care program: Lessons learned though a theory-oriented approach. *Evaluation and Program Planning*, 20, 217-224.
- Rogers, P., Hacsí, T., Petrosino, A., and Huebner, T. (2000). Program Theory in Evaluation: Challenges and Opportunities. *New Directions for Evaluation*, 87, San Francisco, CA: Jossey-Bass.
- Rogers, P., Petrosino, A., Huebner, T. and Hacsí, T. (2000). Program theory evaluation: Practice, promise, and problems. In P. Rogers, T. Hacsí, A. Petrosino, and T. Huebner (Eds.), *New Directions for Evaluation*, 87, 5-13.
- Rossi, P. (1990). Foreword. *Theory Driven Evaluations*. By Huey-Tsyh Chen. Newbury Park, CA: Sage.
- Rossi, P. and Wright, J. (1986). Evaluation research: An assessment. In D. Cordray and M. Lipsey (Eds.), *Evaluation Studies Review Annual*, 11, 48-69.
- Rossi, P., Freeman, H., and Lipsey, M. (1999). *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage Publications.
- Sanders, J. (1990). Curriculum Evaluation Research. In H.J Walberg and G.D. Haertel (Eds.), *The International Encyclopedia of International Evaluation*. Pergamon Press: Oxford.
- Sandler, I., West, S., Baca, L., Pillow, D., Gersten, J., Rogosch, F., Viridin, L., Beals, J., Reynolds, K., Kallgren, C., Tein, J., Kriege, G., Cole, E., and Ramirez, R. (1992). Linking empirically based theory and evaluation: The family bereavement program. *American Journal of Community Psychology*, 20 (4), 491-521.
- Schein, E. (1992). *Organisational Culture and Leadership* (2<sup>nd</sup> ed.), San Francisco: Jossey-Bass.
- Schereir, M. (1987). Program theory and implementation theory: Implications for Evaluators. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 33, 5-18.
- Scriven, M. (1967). The methodology of evaluation. In Robert Stake (Ed.), *Perspectives of Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation*, 1, 39-83. Chicago: Rand McNally.
- Scriven, M. (1971). Evaluating educational programs. In F.G. Caro (Ed.), *Readings in evaluation research* (49-53). New York: Russell Sage Foundation.
- Scriven, M. (1991). *Evaluation Thesaurus* (4<sup>th</sup> ed.). Newbury Park, CA: Sage.
- Scriven, M. (1994). The fine line between evaluation and explanation. *Evaluation Practice*, 15, 75-77.
- Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation*, 19 (1), 57-70.

- Scriven, R. (1967). The methodology of evaluation. In Robert Stake (Ed.), *Perspectives of Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation*, 1, 39-83. Chicago: Rand McNally.
- Seigel, K. and Tuckel, P. (1985). The utilisation of evaluation research: A case analysis. *Evaluation Review*, 9 (3), 307-328.
- Senge, P. (1990). *The Fifth Discipline*, New York: Doubleday Currency.
- Senge, P. (1993). Transforming the practice of management. *Human Resource Development Quarterly*, Spring, 3-11.
- Senge, P., Ross, R., Smith, B., Roberts, C., and Kleiner, A. (1994). *The Fifth Discipline Fieldbook*. London: Nicholas Brealey Publishing Limited.
- Shadish, W. (1987). Program micro- and macrotheories: A guide for social change. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 33, 5-18.
- Shadish, W., Cook, T., and Leviton L. (1991). Social program evaluation: Its history, tasks, and theory. In W. Shadish, T. Cook and L. Leviton (Eds.) *Foundations of Program Evaluation: Theories of Practice* (19-35). Newbury Park; Sage.
- Shapiro, J and Blackwell, D. (1987). Large-scale evaluation on a limited budget: The partnership experience. In J. Nowakowski (Ed.), *The Client Perspective on Evaluation: New Directions for Program Evaluation*, 36, San Francisco: Jossey-Bass.
- Shea, M., Lewko, J., Boschen, K., Flynn, R., and Volpe, R. (1995). Facilitating instrumental utilisation for policy development in a multisite, interministerially sponsored human service program. *The Canadian Journal of Program Evaluation*, 10 (2), 91-106.
- Shulha, L., and Cousins, B. (1995). *Utilisation and social justice: Interconnections of meta-evaluation frameworks*. Paper presented at the joint meeting of the Canadian Evaluation Society and the American Evaluation Association, Vancouver, BC.
- Shulha, L. and Cousins, B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18, 195-208.
- Sidani, S. and Sechrest, L. (1999). Putting program theory into operation. *American Journal of Evaluation*, 20 (2), 227-238.
- Smith, N. (1994). Clarifying and expanding the application of program theory-driven evaluations. *Evaluation Practice*, 15, 83-87.
- Smylie, M. and Wang, M. (1988). Teacher development and school improvement: The process of teacher change. *Teaching and Teacher Education*, 4(2), 171-187.
- Smylie, M. (1988). The enhancement function of staff development: Organisational and psychological antecedents to individual teacher change. *American Educational Research*, 25(1), 1-30.
- Stake, R. and Easley, J. (1978). *Case studies in science education*. Champaign: University of Illinois, Center for Instructional Research and Curriculum Evaluation

- Stake, R. (1967). The countenance of educational evaluation. *Teacher's College Record*, 68 (7), 523-540.
- Stake, R. (1978). The case study method in social inquiry. *Educational Researcher*, 7, 5-8.
- Stake, R. (1991). Retrospective on "The countenance of educational evaluation". In Milbery McLaughlin and D.C. Phillips (Eds.), *Evaluation and Education: At the Quarter Century. Ninetieth Yearbook of the National Society for the Study of Education*, Part 2, Chicago: University of Chicago Press.
- Stamps, D. (1997). Communities of practice. Learning is social, Training is irrelevant? *Training*, 34(2), 34-42.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Straton, R. (1977). *Research of the Evaluation Process: Current Status and Future Directions*. Paper presented at the Annual Conference of the Australian Association for Research in Education, Canberra.
- Straton, R. (1990). Research of the Evaluation Process in Australia. *Evaluation Journal of Australasia*, 2 (2), 2-12.
- Stufflebeam, D. (1980). 'An interview with Daniel L. Stufflebeam' *Educational Evaluation and Policy Analysis*, 2, 4
- Stufflebeam, D. (2000). Lessons in contracting for evaluations. *The American Journal of Evaluation*, 21, 293-314.
- Stufflebeam, D. (2001). Evaluation Models. *New Directions for Evaluation*, 89.
- Stufflebeam, D., Foley, W., Gephart, W. Guba E., Hammond R., Merriman, H., and Provus, M. (1971). *Educational Evaluation and Decision-making*. Itasca, Illinois: Peacock Publishers.
- Suchman, E. (1967). *Evaluative Research: Principles and Practice in Public Services and Social Action Programs*. New York: Russell Sage Foundation.
- Sussman, S., Galaif, E., Newman, T., Hennesy, M., Pentz, M., Dent, C., Stacy, A., Moss, M., Craig, S., and Simon, T. (1997). Implementation and process evaluation of a student "School- As-Community" group: A component of a school-based drug abuse prevention program. *Evaluation Review*, 21 (1), 94-123.
- Tanaka, J. (1987). "How big is big enough?" Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, 58, 134-146.
- Tanaka, J. and Huba, G. J. (1989). A general coefficient of determination for co-variance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 42, 233-239.
- Tanaka, J. and Huba, G. J. (1985). A fit index for co-variance structural models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 42, 233-239.

- Tilley, N. (1996). Demonstration, exemplification, duplication and replication in evaluation research. *Evaluation*, 2 (1), 35-50.
- Torvatn, H. (1999). Using program theory models in evaluation of industrial modernization program: three case studies. *Evaluation and Program Planning*, 22(2), 73-82.
- Trochim, W. (1985). Pattern matching, validity, and conceptualisation in program evaluation. *Evaluation Review*, 9, 575-604.
- Trochim, W. (1989). Outcome pattern matching and program theory. *Evaluation and Program Planning*, 12, 355-366.
- Trochim, W. (1994). Using concept mapping to develop a conceptual framework of staff's views of a supported employment program for persons with severe mental illness. *Journal of Consulting and Clinical Psychology*, 62 (4), 766-775.
- Trochim, W., Cook, J., and Setze, R. (1994). Using concept mapping to develop a conceptual framework of staff's views of a supported employment program for persons with severe mental illness. *Journal of Consulting and Clinical Psychology*, 62 (4), 766-775.
- Tucker, L. R., and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- Turner, G. and Shepherd, J. (1999). A method in search of a theory: peer education and health promotion. *Health Education Research*, 14 (2), 2, 235-247.
- Vlahov, L. (1990). *The Influence of Personal Factors on the Use of Evaluation Information by Upper and Middle Level Administrators*. Unpublished Doctor of Philosophy dissertation, Murdoch University, Western Australia.
- Watkins, K. and Marsick, V. (1992). Towards a theory of informal and incidental learning in organisations. *International Journal of Lifelong Education*, 11, 287-300.
- Webb, J., Zhang, C. and Sillitoe, J. (1999). Improving access to academic support in higher education. *Evaluation News and Comment*, 8 (2), 12-16.
- Weiss, C. and Bucuvalas, M. (1972). The challenge of social research to decision making. In Carol H. Weiss (Ed.), *Evaluating Action Programs in Social Action and Education*. Boston: Allyn and Bacon, Inc.
- Weiss, C. (1972a). *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.
- Weiss, C. (1972b). Evaluating action programs; Readings in social action and education. Boston: Allyn and Bacon.
- Weiss, C. (1972c). Utilisation of evaluation: Toward comparative study. In Carol H. Weiss (Ed.), *Evaluating Action Programs: Readings in Social Action and Education*. Boston: Allyn and Bacon. (Reprint of paper presented at the meeting of the American Sociological Association, Miami Beach, Florida, September, 1966.).



- Weiss, C. (1981). Measuring the use of evaluation. In J.A. Ciarlo (Ed.), *Utilising evaluation: Concepts and measurement techniques*. Sage Research Progress Series in Evaluation, 6, 17-33.
- Weiss, C. (1987). Evaluating social programs: What have we learned? *Society*, 25 (1), 40-45.
- Weiss, C. (1997a). How can theory-based evaluation make greater headway? *Evaluation Review*, 21, 501-524.
- Weiss, C. (1997b). Theory-based evaluation: Past, present, and future. In D. Rog and D. Fournier (Eds.), *New Directions for Program Evaluation*, 76, 41-55.
- Weiss, C. (1998). *Evaluation: Methods for Studying Programs and Policies*. (2<sup>nd</sup> Ed). Englewood Cliffs, N.J.: Prentice Hall, 1998.
- Weiss, C. and Birckmayer, J. (2000). Theory-Based Evaluation in Practice: What Do We Learn? *Evaluation Review*, 24 (4), 407-431.
- West, S., Finch, J. and Curran, P. (1995). Structural Equation Models with Nonnormal Variables: Problems and remedies. In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (56-75). Thousand Oaks: Sage.
- Wheaton, B., Muthén, B., Alwin, D., and Summers, G. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological Methodology* (84-136). San Francisco: Jossey-Bass.
- Wholey, J. (1987). Evaluability assessment: Developing program theory. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 33, 5-18.
- Wholey, J. (1979). *Evaluation: Promise and Performance*. Washington: Urban Institute.
- Williams, L. and Holahan, P. (1994). Parsimony-based fit indices for multiple-indicator models: Do they work? *Structural Equation Modelling*, 1, 161-189.
- Wingens, M. (1990). Toward a general utilisation theory: A systems theory reformulation of the two-communities metaphor. *Knowledge: Creation, Diffusion, Utilisation*, 12, 27-42.
- Worthen, B. (1996). Editor's note: The origins of theory-based evaluation, *Evaluation Practice*, 17, 169-171.
- Yin, R. (1992). The role of theory in doing case study research and evaluations. In H. Chen and P. Rossi (Ed.), *Using Theory to Improve Program and Policy Evaluation*. Westport, CT: Greenwood.
- YOHFest (2004). *Information Pack for Teachers*. Perth, Western Australia.
- Youtie, J., Bozeman, B. and Shapira, P. (1999). Using an evaluability assessment to select methods for evaluating state technology development programs: the case of the Georgia Research Alliance. *Evaluation and Program Planning*, 22(2), 55-64.

Zapelli, R., Hepworth, H., and Morgan, P. (2004, September). *Evaluation of the WA Pre-Apprenticeship Program: A Research Report*. Perth, Western Australia.

## **APPENDIX A**

1. Information Letter for Interviewees
2. Consent Form to be signed by Interviewee
3. Organisational Environment Characteristics Interview Schedule

Dear \_\_\_\_\_

I look forward to including your responses to my interview questions in my study data.

As you know, I am a Ph.D. student at Murdoch University researching the use of evaluations and their information. This current study extends research completed as part of my Master's Degree considering the influences on the use of evaluation information such as the nature of the host organisation, the stakeholders and the evaluation process.

All involvement I have with any program, its staff and evaluation will be conducted in accordance with the ethical guidelines for practice advocated by the Australasian Evaluation Society, and the Murdoch University research ethics code of practice regarding human research. I will undertake precautions to ensure no names or other information that may identify you are reported in any informal paper or publication linked to this study. All completed interview schedules will be coded and stored in a secure area.

I will make a summary of the study findings available to you when I have finished. Clearly, sound research considering the evaluation process and the factors influencing that process should provide insightful information useful to the organisation funding the evaluation and the program evaluated.

Should you think of any questions you would like to ask me regarding this research at a later time my contact details are below.

Kind Regards,

Sheri Hudson-Mabbs

**MURDOCH UNIVERSITY**

*24 Aberfoyle Heights*

*Currambine WA 6028*

*(08) 9305 3847*

*[bigskies@iprimus.com.au](mailto:bigskies@iprimus.com.au)*

## RESEARCH CONSENT FORM

**Project Title: Factors Influencing the Use of Evaluation Information: A Longitudinal Study**

I am a Ph.D. student at Murdoch University investigating the influence of factors on the use of program evaluation information in organisations. The purpose of the study is to identify more efficient and useful evaluation practices and processes.

You can help in this study by consenting to be interviewed. Contained in the interview schedule are questions that relate to the evaluation of the XXXXXX program, and the use of its information. You may decide to withdraw your consent at any time.

All information given during the interview will be confidential. No names or other information that may identify you will be used in any publication arising from the study. Feedback on the study will be provided to you.

If you are willing to participate in this study, could you please complete the details below? If you have any questions about this project please feel free to contact either myself, Sheri Hudson-Mabbs, on 9305 3847, Associate Professor Ralph Straton on 9360 6995, or Dr. Richard Cummings on 9360 2354.

My supervisors and I are happy to discuss with you any concerns you may have on how this study is conducted, or you can contact Murdoch University's Human Research Ethics Committee on 9360 6677.

\*\*\*\*\*

I (the participant) have read the information above. Any questions I have asked have been answered to my satisfaction. I agree to take part in this activity; however I may change my mind and withdraw at any time.

I understand that all information provided will be treated as confidential and will not be released by the investigator unless required to do so by law.

I agree for this interview to be taped.

I agree that research data gathered for this study may be published provided my name or other information which might identify me is not used.

Participant/Authorized Representative's Name:	Investigator's Name:
Participant/ Authorized Representative's Signature:	Investigator's Signature:
Date:	Date:

**Code:** \_\_\_\_\_

## **Organisational Environment Characteristics Interview Schedule**

**Sheri Hudson-Mabbs**

**The Australian Institute of Education**

**Murdoch University**

<b>Name:</b>
<b>Date:</b>
<b>Title:</b>
<b>Program:</b>
<b>Organisation:</b>

## **Introduction**

The purpose of my research is to investigate the influence of program evaluations. One factor found to influence evaluation effect is the environment of the organisation responsible for the delivery of the evaluated program. The questions included in the following pages have been developed to gather information from you about the environment of the organisation within which you work and your personal practice.

For each of the items below, respond with the number that best represents your opinion based on your experiences, and not on how you think other individuals would answer, or your organisation's official policy or intent. In instances where you are not certain about a response please select the response you consider the most likely. If you have any questions please don't hesitate to ask me. It is important to the quality of the information gathered in this interview that we have a shared understanding of both my questions and your responses.

The first section of this interview asks about the general environment of your organisation. The second and third sections include questions about the organisation's managerial or supervisory staff and your personal approach to your work in the organisation, respectively. In the fourth section, the questions focus on the use of information in the organisation. The final section, which concentrates on teamwork in the organisation, is followed by a short series of questions about you.



## I. Organisation Environment

Could you please consider the organisation chart I have here and locate the division or section you identify with. Now, when responding to the following questions could you please consider them in terms of the division or section you identify with in your organisation. The questions in this section ask you about your views regarding the environment of the organisation you work for. 'Organisation Environment' is defined here as the general practices and behaviour of the organisation. Please respond with a rating to each question on a scale of 1 to 5, 1 representing 'Not at All' and 5 representing 'A Great Deal' (Give respondent an example of the scale).

To what extent:		Not at all      A Great Deal				
		1	2	3	4	5
1.	Is there a clear understanding of what the Organisation is seeking to achieve?	1	2	3	4	5
2.	Are the expectations of the organisation regarding staff workload realistic?	1	2	3	4	5
3.	Does the organisation regularly look for ways to improve policies, programs or services?	1	2	3	4	5
4.	Does the organisation support the learning of new knowledge and skills by the staff?	1	2	3	4	5
5.	Is open dialogue regarding organisation policies or practices encouraged?	1	2	3	4	5
6.	Is negative feedback regarding the policies or practices of the organisation encouraged?	1	2	3	4	5
7.	Are all assumptions, policies or practices open for discussion?	1	2	3	4	5
8.	Is competition encouraged among the branches or units of the organisation?	1	2	3	4	5
9.	Is collaboration among the organisation's branches or units encouraged?	1	2	3	4	5
10.	Does the organisation support the trialling of new knowledge and skills by the staff?	1	2	3	4	5
11.	a. Are unsuccessful organisational policies, strategies, programs, etc., considered primarily in terms of (please choose only one): knowledge gained from the experience?      or resources lost in the experience?					<input type="checkbox"/> <input type="checkbox"/>
	b. To what extent are unsuccessful undertakings considered in terms of (REPEAT CHOICE FROM 'B' TO RESPONDENT)?	1	2	3	4	5

## II. Leadership Practice

This section focuses on your views regarding the approach of those employed by the organisation in a managerial or supervisory capacity. Once again, please respond with a rating to each question on a scale of 1 to 5, 1 representing 'Not at All' and 5 representing 'A Great Deal'.

To what extent:		Not at all		A Great Deal		
		1	2	3	4	5
12.	Do managerial staff have a clear understanding of what the organisation is seeking to achieve?	1	2	3	4	5
13.	Do managerial staff make realistic work commitments for themselves?	1	2	3	4	5
14.	Do managerial staff regularly look for ways to improve policies, processes, programs, or services?	1	2	3	4	5
15.	Do managerial staff model the importance of learning through their own efforts to learn?	1	2	3	4	5
16.	Do managerial staff openly talk about organisation policies or practices?	1	2	3	4	5
17.	Do they offer constructive feedback regarding the policies or practices of the organisation?	1	2	3	4	5
18.	Are there any assumption, practices, policies, etc., that managers refrain from offering an opinion or viewpoint on?	1	2	3	4	5
19.	Are managerial staff open to constructive feedback from staff?	1	2	3	4	5
20.	Do managerial staff encourage staff to work together?	1	2	3	4	5
21.	Do managerial staff encourage those they supervise to compete with each other?	1	2	3	4	5
22.	Are managerial staff supportive of trial programs, policies and strategies based on new knowledge?	1	2	3	4	5
23.	a. Do managerial staff consider unsuccessful organisational policies, strategies, programs, etc., primarily in terms of (please choose only one):					
	knowledge gained from the experience? or					<input type="checkbox"/>
	resources lost in the experience?					<input type="checkbox"/>
	b. To what extent are unsuccessful undertakings considered in terms of (REPEAT ABOVE CHOICE TO RESPONDENT)?	1	2	3	4	5

### III. Personal Practice

The questions in this section ask about how you manage and approach your work at the organisation. Again, please respond on the same five point scale.

To what extent:		Not at all      A Great Deal				
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
24.	Do you have a clear understanding of what the organisation is seeking to achieve?	1	2	3	4	5
25.	Do you make realistic work commitments for yourself?	1	2	3	4	5
26.	Do you regularly look for ways to improve policies, processes, programs, or services?	1	2	3	4	5
27.	Do you pass up opportunities to learn new knowledge and skills applicable to work?	1	2	3	4	5
How come?						
28.	Do you openly talk about organisation policies or practices?	1	2	3	4	5
29.	Do you give negative feedback regarding organisation policy or practice when you feel it is necessary?	1	2	3	4	5
30.	Are there any assumptions, practices, etc., that you refrain from offering an opinion or viewpoint on?	1	2	3	4	5
31.	Do you encourage negative feedback from your colleagues at work?	1	2	3	4	5
32.	Do you cooperate with your workmates over work issues?	1	2	3	4	5
33.	Are you competitive with your workmates over work issues?	1	2	3	4	5
34.	Do you take opportunities to trial programs, policies or strategies based on new knowledge?	1	2	3	4	5
35.	a. Do you consider unsuccessful organisational policies, strategies, programs, etc., primarily in terms of (please choose only one):					
	knowledge gained from the experience? or					
	resources lost in the experience?					
	b. To what extent are unsuccessful undertakings considered in terms of					
	(REPEAT ABOVE CHOICE TO RESPONDENT)?	1	2	3	4	5

#### IV. Use of Information

The questions in this section concentrate on your views about the information collection, dissemination and use routines of the organisation. Once again, please respond on the five point scale.

To what extent:		Not at all		A Great Deal		
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
36.	Is information regularly gathered from clients and other stakeholders to gauge how well programs and activities are doing?	1	2	3	4	5
37.	Are there adequate records of past change efforts and what happened as a result?	1	2	3	4	5
38.	Are formal evaluations of the organisation's activities and programs routinely undertaken? (An evaluation is the formal activity of investigating the program, activity, etc. and reporting the results to decision-makers.)	1	2	3	4	5
39.	Are formalised evaluation activities incorporated into the delivery of organisational programs?	1	2	3	4	5
40.	Does the organisation collect information it needs?	1	2	3	4	5
41.	Are systems adequate to disseminate information gathered by the organisation to those staff who need and can use it?	1	2	3	4	5
42.	Is it difficult for staff to access information collected by the organisation to make decisions regarding their work?	1	2	3	4	5
43.	Do staff use the information collected by the organisation in their work practice?	1	2	3	4	5
44.	Do staff rely on the information collected by the organisation in the decisions they make regarding their work?	1	2	3	4	5
45.	Do managerial staff tend to use information collected by the organisation in their decisions?	1	2	3	4	5

To what extent:	Not at all		A Great Deal		
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
46. To what extent does the organisation alert staff to new information?	1	2	3	4	5
Why or why not?					
47. Does the organisation encourage staff to share information?	1	2	3	4	5
Why or why not?					

## V. Teamwork

This section asks a few questions about the use of teamwork in your organisation and the team approach.

To what extent:

48. Is your work sometimes conducted as part of a working group that is or could be identified as a “team”?
- ☐ Yes (Continue with item 45)
- ☐ No (Go to item 54)

Once again on the five point scale, 1 representing ‘Not at All’ and 5 representing ‘A Great Deal’,

To what extent:		Not at all      A Great Deal				
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
49.	Are employees provided training on how to work as a team member?	1	2	3	4	5
50.	Are team meetings well attended?	1	2	3	4	5
51.	Is task-oriented dialogue a part of team meetings?	1	2	3	4	5
52.	Is an effort made at team meetings to develop a vision of the team tasks and goals that are <i>shared</i> by all team members?	1	2	3	4	5
53.	Is conflict that arises among team members resolved effectively?	1	2	3	4	5
54.	Do team members go along with decisions they don't really agree with?	1	2	3	4	5
55.	Are teams more productive than individuals working alone?	1	2	3	4	5
56.	Do teams fail to accomplish work they are charged to do?	1	2	3	4	5
57.	Is knowledge developed in teams shared with other groups or co-workers?	1	2	3	4	5

## VI. Some Final Questions

**Lastly, a few questions about yourself.**

58. Which category best describes your position at the organisation?

- ☐ Executive
- ☐ Administrative
- ☐ Supervisory/ Managerial
- ☐ Service/Program Practitioner

59. How many years have you been employed in your present position? \_\_\_\_\_years

60. How many years have you worked for this organisation? \_\_\_\_\_years

61. How long do you intend to stay with this organisation?

- ☐ 0 – 2 years
- ☐ 3 – 5 years
- ☐ 6 – 10 years
- ☐ 11 – 15 years
- ☐ 16 – 20 years
- ☐ 21 years or more

59. Do you have anything you would like to add before we close?

## **APPENDIX B**

1. Information Letter for Interviewees
2. Consent Form to be signed by Interviewee
3. Stakeholder Interview Schedule I
4. Process Use Interview Schedule
5. Stakeholder Interview Schedule II



Dear \_\_\_\_\_

I look forward to including your responses to my interview questions in my study data.

As you know, I am a Ph.D. student at Murdoch University researching the use of evaluations and their information. This current study extends research completed as part of my Master's Degree considering the influences on the use of evaluation information such as the nature of the host organisation, the stakeholders and the evaluation process.

All involvement I have with any program, its staff and evaluation will be conducted in accordance with the ethical guidelines for practice advocated by the Australasian Evaluation Society, and the Murdoch University research ethics code of practice regarding human research. I will undertake precautions to ensure no names or other information that may identify you are reported in any informal paper or publication linked to this study. All completed interview schedules will be coded and stored in a secure area.

I will make a summary of the study findings available to you when I have finished. Clearly, sound research considering the evaluation process and the factors influencing that process should provide insightful information useful to the organisation funding the evaluation and the program evaluated.

Should you think of any questions you would like to ask me regarding this research at a later time my contact details are below.

Kind Regards,

Sheri Hudson-Mabbs  
**MURDOCH UNIVERSITY**  
*24 Aberfoyle Heights*  
*Currambine WA 6028*  
*(08) 9305 3847*  
[bigskies@iprimus.com.au](mailto:bigskies@iprimus.com.au)

## RESEARCH CONSENT FORM

**Project Title: Factors Influencing the Use of Evaluation Information: A Longitudinal Study**

I am a Ph.D. student at Murdoch University investigating the influence of factors on the use of program evaluation information in organisations. The purpose of the study is to identify more efficient and useful evaluation practices and processes.

You can help in this study by consenting to be interviewed. Contained in the interview schedule are questions that relate to the evaluation of the XXXXXX program, and the use of its information. You may decide to withdraw your consent at any time.

All information given during the interview will be confidential. No names or other information that may identify you will be used in any publication arising from the study. Feedback on the study will be provided to you.

If you are willing to participate in this study, could you please complete the details below? If you have any questions about this project please feel free to contact either myself, Sheri Hudson-Mabbs, on 9305 3847, Associate Professor Ralph Straton on 9360 6995, or Dr. Richard Cummings on 9360 2354.

My supervisors and I are happy to discuss with you any concerns you may have on how this study is conducted, or you can contact Murdoch University's Human Research Ethics Committee on 9360 6677.

\*\*\*\*\*

I (the participant) have read the information above. Any questions I have asked have been answered to my satisfaction. I agree to take part in this activity; however I may change my mind and withdraw at any time.

I understand that all information provided will be treated as confidential and will not be released by the investigator unless required to do so by law.

I agree for this interview to be taped.

I agree that research data gathered for this study may be published provided my name or other information which might identify me is not used.

Participant/Authorized Representative's Name:	Investigator's Name:
Participant/ Authorized Representative's Signature:	Investigator's Signature:
Date:	Date:

Code: \_\_\_\_\_

### Stakeholder Interview Schedule I

**Sheri Hudson-Mabbs  
Murdoch University  
October 1, 2009**

<b>Name:</b>
<b>Date:</b>
<b>Title:</b>
<b>Program:</b>
<b>Organisation:</b>

## **Introduction**

The purpose of this questionnaire is to gather some information from the stakeholders of the \_\_\_\_\_ program. For the purpose of this questionnaire a stakeholder is considered to be someone with an interest in the program. I will be asking you questions about your involvement with the program, your experience with social research and evaluation, current information use in the program and the extent to which you expect to be involved in the evaluation of the program. Although a few of the questions require short answer responses most items call for an answer rated on a 5 point scale. Your responses to this questionnaire are appreciated.

### **I. The \_\_\_\_\_ Program**

The questions included in this section focus on your views of the program.

1. How many years have you been (or were you) involved with the program?
  
  
  
  
  
  
  
  
  
  
2. How have you been involved with the program?

**Now I'm going to ask you some questions about your involvement with the program and about your views of the program. Please answer the following questions on the scale provided.**

To what extent:		Not at all		A Great Deal		
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
3.	Were you involved with the early development of the program?	1	2	3	4	5
4.	Were you involved with the initial delivery of the program?	1	2	3	4	5
5.	Are you currently involved in the delivery of the program?	1	2	3	4	5
6.	Do you feel there is a need for the program?	1	2	3	4	5
7.	Do you feel the intended outcomes of the program are worthwhile?	1	2	3	4	5
8.	Do you agree with the manner in which the program services/ practices are delivered?	1	2	3	4	5
9.	Do you feel there are better ways of delivering this program?	1	2	3	4	5
10.	Do you feel the program is based on a sound philosophy?	1	2	3	4	5

The next three questions are about your influence on the development, delivery and outcomes of the program.

To what extent:		Not at all		A Great Deal		
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
11.	Do you feel you have personally influenced the <b>development</b> of this program?	1	2	3	4	5
12.	Do you feel you personally influence the <b>delivery</b> of this program?	1	2	3	4	5
13.	Do you feel you personally influence the <b>outcomes</b> of this program?	1	2	3	4	5

Finally, to finish this section...

14.	Overall, do you support or oppose the program?	Support				
		Oppose				
15.	How strongly do you support/ oppose the program?	1	2	3	4	5

## II. The Evaluation Study

**Now, I would like to ask you some questions about the proposed evaluation study.**

16. What do you see as the main purpose(s) of the proposed evaluation study?

**Please answer the following questions on the scale provided.**

To what extent:	Not at all		A Great Deal		
	1	2	3	4	5
17. Do you feel there is a need for this evaluation study?	1	2	3	4	5
18. Do you expect this study will provide information useful to the program?	1	2	3	4	5
19. Do you expect this study will provide information useful to _____ (the organisation responsible for the program)?	1	2	3	4	5
20. Is the program at a stage in it's life where an evaluation is appropriate? _____DK/NA	1	2	3	4	5
21. Overall, do you support or oppose this evaluation study?	Support		Oppose		
22. How strongly do you support/oppose this study?	1	2	3	4	5
23. Do you believe the findings of the evaluation study will generally be supportive or critical of the program?	Support		Critical		
24. How so?					
25. Do you think the findings will be supportive/critical of the program?	1	2	3	4	5

### III. Involvement Expectation

Evaluation Studies usually have the following three stages- planning the study, gathering and analysing the information and reporting the results. I'd like to ask you some questions about your expectations regarding involvement in each of these stages.

To what extent:		Not at all		A Great Deal		
		1	2	3	4	5
26.	Would you <u>like</u> to be involved in <u>planning</u> the evaluation study?	1	2	3	4	5
27.	Will you be <u>expected</u> to be involved in planning the evaluation study (e.g. by your own organisation, the organisation responsible for the program)?	1	2	3	4	5
28.	Would you <u>like</u> to be involved in the <u>conduct</u> of the evaluation study?	1	2	3	4	5
29.	Will your involvement in the conduct of the evaluation study be <u>expected</u> ?	1	2	3	4	5
30.	Would you <u>like</u> the evaluation team to report regularly?	1	2	3	4	5
31.	Would you <u>like</u> to be in <u>regular communication</u> with the evaluation team?	1	2	3	4	5
32.	Would you <u>like</u> to be <u>interviewed</u> as part of the study?	1	2	3	4	5
33.	Would you <u>like</u> to be able to <u>feedback</u> to the evaluation team on issues and concerns regarding the evaluation?	1	2	3	4	5
34.	Would you like to be involved in any work to understand or explain the logic, rationale or theory underpinning the program? <b><i>By the logic, rationale or theory of the program I mean the underlying assumptions about how a program will work to achieve intended outcomes.</i></b>	1	2	3	4	5
35.	Would you <u>like</u> to receive the evaluation findings as they emerge?	1	2	3	4	5
36.	Would you <u>like</u> to be involved in <u>developing the recommendations</u> from the study?	1	2	3	4	5



#### IV. Training and Experience with Research and Evaluation

**Now, I would like to ask you a few questions about your background with social research and evaluation. First I'd like to ask you about social research.**

37. To what extent has your educational background included some exposure to social research? 1 2 3 4 5  
**If yes**

38. Can you summarise any education you have undertaken that focused on social research methods. (e.g. workshops, courses, higher education, etc...).

39. To what extent has your training or experience included some exposure to social research? 1 2 3 4 5

40. Can you summarise any other training or experience you've had with social research methods?

41. If any, how many research projects have you been involved with? \_\_\_\_\_

42. How were you involved in these projects?

#### IV. Training and Experience with Research and Evaluation

Please answer the following questions on the 5 pt scale provided.

To what extent:		Not at all		A Great Deal		
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
43.	To what extent have you been able to read social research reports or articles?	1	2	3	4	5
44.	To what extent are you able to consider the implications of research study information for your work?	1	2	3	4	5
45.	To what extent are you able to apply research study information to your work?	1	2	3	4	5

#### IV. Training and Experience with Research and Evaluation

I've just been asking you about social research. Now, I'd like to ask you a few questions about evaluation.

To what extent:	Not at all		A Great Deal		
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>

- |     |                                                                                     |   |   |   |   |   |
|-----|-------------------------------------------------------------------------------------|---|---|---|---|---|
| 46. | To what extent has your educational background included any exposure to evaluation? | 1 | 2 | 3 | 4 | 5 |
|-----|-------------------------------------------------------------------------------------|---|---|---|---|---|

**If yes**

47. Please summarise any education you have undertaken that included evaluation training (e.g. workshops, courses, higher education, etc...).

- |     |                                                                                      |   |   |   |   |   |
|-----|--------------------------------------------------------------------------------------|---|---|---|---|---|
| 48. | To what extent has your training or experience included some exposure to evaluation? | 1 | 2 | 3 | 4 | 5 |
|-----|--------------------------------------------------------------------------------------|---|---|---|---|---|

49. Can you summarise any other training or experience you've had with evaluation methods?

50. If any, how many evaluation projects have you previously been involved with? \_\_\_\_\_

**If yes**

51. How were you involved in these projects?

#### IV. Training and Experience with Research and Evaluation

Please answer the following questions on the scale provided.

	Not at all		A Great Deal		
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
52. To what extent have you been able to read <b>evaluation</b> reports or articles?	1	2	3	4	5
53. To what extent have you been able to consider the implications of <b>evaluation</b> information for your work?	1	2	3	4	5
54. To what extent have you been able to apply <b>evaluation</b> information to your work?	1	2	3	4	5
55. To what extent do you believe <b>evaluation</b> helps us provide better programs, processes, and services?	1	2	3	4	5

V. Views of the \_\_\_\_\_  
(organisation responsible for program delivery).

Next, would like to ask you a few questions about the organisation.

To what extent:		Not at all      A Great Deal				
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
56.	Do you agree with the main strategic direction of the organisation? _____DK/NA	1	2	3	4	5
57.	Do you feel there is a need for the services/ programs of this organisation? _____DK/NA	1	2	3	4	5
58.	Do you feel there are better ways of delivering organisation services? _____DK/NA	1	2	3	4	5
59.	Do you feel the organisation has a positive culture? _____DK/NA	1	2	3	4	5
60.	Do you believe the organisation has the capacity to change if more effective methods of program/service delivery are identified? _____DK/NA	1	2	3	4	5
61.	Do you feel your role in relation to the organisation is effective? _____DK/NA	1	2	3	4	5

## **VI. Use of Information**

**The questions below focus on your experience with the use of program information, but only in relation to this program.**

62. When was the last formal evaluation undertaken?

63. Have formal evaluations of the program been undertaken recently?

64. On the 5 pt scale, to what extent are formalized evaluation activities incorporated into the delivery of the program? 1 2 3 4 5

How so?

65. Again, on the 5 pt scale, to what extent do you rely on information collected about the program in making relevant decisions about the program? 1 2 3 4 5

How so?

66. What sort of information do you refer to?

67. Are there any other comments you would like to make before we close?

**Thank-you for your time and for your responses**

**STAKEHOLDER'S PROCESS USE  
INTERVIEW SCHEDULE**

**Sheri Hudson-Mabbs  
Murdoch University  
October 1, 2009**

## Process Interview

**The purpose of this short interview is to gather some information about your involvement in the evaluation process.**

1. What contact have you had with the evaluators?

---

---

---

---

---

---

---

---

2. What discussions have you have with the evaluators in which you provided program information for the evaluation study?

---

---

---

---

---

---

---

---



3. What discussions have you had with the evaluators in which you had input into the planning or management of the evaluation?

---

---

---

---

---

---

---

---

4. What groups have you been involved with which have been developed for the evaluation study process?

---

---

---

---

---

---

---

---

5. What other evaluation study related activities have you been involved with?

---

---

---

---

---

---

---

---

6. In what ways did your involvement affect your thinking about the program?

---

---

---

---

---

---

---

---

	<u>Not at All</u>		<u>A Great Deal</u>	
7. On a 5 point scale, <b>1</b> representing <b>Not At All</b> and <b>5</b> representing <b>A Great Deal</b> , to what extent do you think this involvement affected your thinking about the program?	1	2	3	4 5

8. In what ways did your involvement affect your thinking about issues other than the program?

-----

-----

-----

-----

-----

-----

-----

-----

-----

	<u>Not at All</u>		<u>A Great Deal</u>	
9. Back on the 5 point scale, to what extent do you think your involvement affected your thinking about issues other than the program?	1	2	3	4 5

10. In what ways did your involvement influence any decisions you were involved in making?

-----

-----

-----

-----

-----

-----

-----

-----

	<u>Not at All</u>			<u>A Great Deal</u>
11. Again on the 5 point scale, to what extent do you think your involvement influenced any decisions you were involved in making?	1	2	3	4 5

12. In what ways did your involvement influence your practice?

-----

-----

-----

-----

-----

-----

-----

13. Once more on the 5 point scale, to what extent do you think your involvement influenced your practice?	1	2	3	4	5
------------------------------------------------------------------------------------------------------------	---	---	---	---	---

14. Are there any further comments you would like to make?

-----

-----

-----

-----

-----

-----

-----

**Thank-you for your time and for your responses**

## **Stakeholder Interview Schedule II**

**Sheri Hudson-Mabbs  
Murdoch University  
October 1, 2009**

The purpose of this questionnaire is to gather some information from the stakeholders of the \_\_\_\_\_ program. For the purpose of this questionnaire a stakeholder is considered to be someone with an interest in the program. I will be asking you questions about your involvement with the program, your experience with social research and evaluation, current information use in the program and the extent to which you expect to be involved in the evaluation of the program. Although a few of the questionnaire items require short answer responses most items call for an answer rated on a 5 point scale. Your responses to this questionnaire are appreciated.

## I. The Evaluation Study

First, I would like to ask you some questions about the \_\_\_\_\_ evaluation study.

1. What do you think were the main purpose(s) of the evaluation study?

Please answer the following questions on the scale provided.

To what extent:

Not at all      A Great  
Deal  
1   2   3   4   5

2. Do you feel this evaluation study was needed?
3. Do you think this study provided information useful to the program?
4. Do you feel the quality of the data that has been collected in the study is sound?      1   2   3   4   5

## I. The Evaluation Study (continued)

To what extent:

- |     |                                                                                                             |                   |   |   |   |   |
|-----|-------------------------------------------------------------------------------------------------------------|-------------------|---|---|---|---|
| 5.  | Do you think this study provided information useful to _____(the organisation responsible for the program)? | 1                 | 2 | 3 | 4 | 5 |
| 6.  | Was the evaluation appropriate for the stage of the program's life?                                         | 1                 | 2 | 3 | 4 | 5 |
| 7.  | Were the evaluation findings timely for any pending decisions?                                              | 1                 | 2 | 3 | 4 | 5 |
| 8.  | Were the implications of the study politically acceptable?                                                  | 1                 | 2 | 3 | 4 | 5 |
| 9.  | Were other sources of information in agreement with the study findings?                                     | 1                 | 2 | 3 | 4 | 5 |
| 10. | Did your organisational role influence the way you used the evaluation information?                         | 1                 | 2 | 3 | 4 | 5 |
| 11. | Do you feel the evaluation study method was appropriate?                                                    | 1                 | 2 | 3 | 4 | 5 |
| 12. | Overall, do you support or oppose this study?                                                               | Support<br>Oppose |   |   |   |   |
| 13. | How strongly do you support/oppose this study?                                                              | 1                 | 2 | 3 | 4 | 5 |
| 14. | To what extent do you believe the findings of the study are supportive of the program?                      | 1                 | 2 | 3 | 4 | 5 |
| 15. | In what ways, if any, was it supportive?                                                                    |                   |   |   |   |   |

**I. The Evaluation Study  
(continued)**

**To what extent:**

- 16.** To what extent do you believe the findings of the study are critical of the program? 1 2 3 4 5
- 17.** In what ways, if any, was it critical?
- 18.** Do the study findings agree with your own personal assessment of the program? 1 2 3 4 5
- 19.** If so, how so?
- 20.** To what extent did the evaluation information influence your thinking about the program? 1 2 3 4 5



## **I. The Evaluation Study (continued)**

**To what extent:**

- 21.** To what extent did the evaluation information influence your thinking about issues beyond the program? 1 2 3 4 5
- 22.** If so, how so?
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 23.** To what extent did the evaluation have an impact on any program decisions? 1 2 3 4 5
- 24.** If so, how so?
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 25.** To what extent did the evaluation have a direct impact on any decisions beyond the immediate program (e.g. guidelines for developing other programs, staffing policy, etc...)? 1 2 3 4 5
- 26.** If so, how so?

## **I. The Evaluation Study (continued)**

**To what extent:**

- 27.** To what extent do you think the evaluation information was used as support for views or positions about the program which people already held? 1 2 3 4 5
- 28.** If so, how so?

- 29.** To what extent do you feel the evaluation focused on issues considered relevant by those interested in the program? 1 2 3 4 5
- 30.** If so, how so?

## II. Evaluators

**In this section I ask you about the conduct of the evaluation study.**

31. Have you met with the evaluator(s)? yes    no
32. On what occasions did you meet (how many times, etc..)?
33. Why do you think the evaluators were chosen to conduct the evaluation (e.g. special skills, background, history, affordability, etc...)?
34. To what extent do you think the evaluators chosen were sufficiently competent to undertake the evaluation study? 1   2   3   4   5
35. To what extent was the evaluation team approachable? 1   2   3   4   5
36. To what extent was the evaluation team protective of their ideas for the evaluation? 1   2   3   4   5
37. To what extent did the evaluation team endeavour to develop relationships with stakeholders? 1   2   3   4   5
38. To what extent did the evaluation team listen to the ideas stakeholders had regarding the evaluation, even when they challenged their own? 1   2   3   4   5

## **II. Evaluators (continued)**

**To what extent:**

39. To what extent were there times when the evaluation process was unable to proceed as planned? 1 2 3 4 5
40. If so, when did these occasions arise and how were they resolved?
41. To what extent were the evaluators tolerant of change? 1 2 3 4 5
42. Why or why not?
43. To what extent do you feel the evaluators were given sufficient latitude to exercise sound professional judgment in undertaking the evaluation? 1 2 3 4 5
44. Why or why not?

## **II. Evaluators (continued)**

45. Considering circumstances surrounding the evaluation, to what extent do you feel the evaluators were free to undertake the evaluation? 1 2 3 4 5
46. Why or why not
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
47. Given those circumstances, to what extent were the evaluators competent at conducting the evaluation? 1 2 3 4 5
48. Why or Why not?

### III. Study Involvement

Evaluation studies usually have the following three stages- planning the study, gathering and analysing the information and reporting the results. I'd like to ask you some questions about your involvement in each of these stages.

**To what extent:**

- |     |                                                                                                                                     |   |     |   |    |   |
|-----|-------------------------------------------------------------------------------------------------------------------------------------|---|-----|---|----|---|
| 49. | Were you involved in planning the evaluation study?                                                                                 | 1 | 2   | 3 | 4  | 5 |
| 50. | Was your involvement in planning the evaluation study expected (e.g. by org. responsible for delivery, your org. ect...)?           | 1 | 2   | 3 | 4  | 5 |
| 51. | Were you involved in the conduct of the evaluation study?                                                                           | 1 | 2   | 3 | 4  | 5 |
| 52. | Was your involvement in the conduct of the evaluation study as expected (e.g. by org. responsible for delivery, your org., ect...)? | 1 | 2   | 3 | 4  | 5 |
| 53. | Were you involved in regular communication with the evaluation team?                                                                | 1 | 2   | 3 | 4  | 5 |
| 54. | Did the evaluation team report regularly?                                                                                           |   | yes |   | no |   |
| 55. | Were you interviewed as part of the study?                                                                                          |   | yes |   | no |   |
| 56. | Did you provide feedback to the evaluation team on issues and concerns regarding the evaluation?                                    | 1 | 2   | 3 | 4  | 5 |
| 57. | Did you receive the evaluation findings as they were reported?                                                                      |   | yes |   | no |   |
| 58. | Were you involved in developing the recommendations from the study?                                                                 | 1 | 2   | 3 | 4  | 5 |

#### IV. The Influence of Program Theory Use

Now I am going to ask you a few questions about the use of the program's theory, logic or rationale. By the program's theory or logic I mean the underlying assumptions about how a program will work to achieve intended outcomes. The questions below focus on the use of the program's theory, logic or rationale in the evaluation and on how that may have influenced your thinking and practice. Some of these questions might require some thought, so please feel free to take your time responding.

To what extent:

59. Did the evaluation report contain information about the program theory or logic? (If '**Not At All**' proceed to next section.) 1 2 3 4 5
60. Was it presented in a manner that made sense? 1 2 3 4 5
61. What do you think was the source of the program's theory used by the evaluator's (prompts: existing program literature, program practitioners, their own observations, etc...)?
62. On a 5 point scale, 1 representing **Not At All** and 5 representing **A Great Deal**, do you agree with the program theory or logic used in the evaluation? 1 2 3 4 5
63. Why or why not?

#### IV. The Influence of Program Theory Use (continued)

64. In what ways do you believe the use of the program's theory in the evaluation influenced your thinking about the program?
65. On a 5 point scale, 1 representing **Not At All** and 5 representing **A Great Deal**, To what extent do you believe the use of the program's theory in the evaluation influenced your thinking about the program? 1 2 3 4 5
66. In what ways do you believe the use of the program's theory in the evaluation influenced your practice?
67. On a 5 point scale, 1 representing **Not At All** and 5 representing **A Great Deal**, To what extent do you believe the use of the program's theory in the evaluation influenced your practice? 1 2 3 4 5



#### IV. The Influence of Program Theory Use (continued)

68. In what ways do you believe the use of the program's theory in the evaluation influenced any decisions you've made or will make about the program?
69. On a 5 point scale, 1 representing **Not At All** and 5 representing **A Great Deal**, to what extent do you believe the use of the program's theory in the evaluation influenced any decisions you've made or will make about the program?
70. On a 5 point scale, 1 representing **Not At All** and 5 representing **A Great Deal**, to what extent were you involved in any work to understand/explain the theory or logic underpinning the program? (If '**Not at All**' continue to next section - Use of Evaluation Information)
- 1 2 3 4 5

#### IV. The Influence of Program Theory Use (continued)

71. If so, how were you involved?
72. In what ways did this focus on program theory influence your thinking about the program?
73. On a 5 point scale, 1 representing **Not At All** and 5 representing **A Great Deal**, To what extent did this focus on program theory influence your thinking about the program?
- 1 2 3 4 5
74. In what ways did this focus on program theory influence your practice?
75. On a 5 point scale, 1 representing **Not At All** and 5 representing **A Great Deal**, to what extent did this focus on program theory influence your practice?
- 1 2 3 4 5

#### IV. The Influence of Program Theory Use (continued)

76. In what ways did this focus on program theory influence any decisions you've made or expect to make about the program?
77. On a 5 point scale, 1 representing **Not At All** and 5 representing **A Great Deal**, to what extent did this focus on program theory influence any decisions you've made or expect to make about the program?

1 2 3 4 5

## V. Use of Evaluation Information

Now I'm going to ask you a few questions about how you have or may use the evaluation report and its information.

- |     |                                                                                                                                                           |                   |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| 78. | <b>Have you been able to read the evaluation report?<br/>If yes, go to Q81</b>                                                                            | <b>yes    no</b>  |
| 79. | How likely is it that you will read the report?                                                                                                           | 1   2   3   4   5 |
| 80. | How important to you is reading the evaluation report?<br><b>Go to Q82</b>                                                                                | 1   2   3   4   5 |
| 81. | How important was it for you to read the evaluation report?                                                                                               | 1   2   3   4   5 |
| 82. | <b>Have you been able to hold informal discussions about the evaluation information with your colleagues? If yes, go to Q85</b>                           | 1   2   3   4   5 |
| 83. | How likely is it you will initiate informal discussions about the information with your colleagues?                                                       | 1   2   3   4   5 |
| 84. | How important to you is discussing the evaluation study information with your colleagues? <b>Go to Q86</b>                                                | 1   2   3   4   5 |
| 85. | How important was discussing the evaluation study information with your colleagues?                                                                       | 1   2   3   4   5 |
| 86. | <b>Have you been able to discuss the evaluation information at a formal meeting? If yes, go to Q89</b>                                                    | <b>yes    no</b>  |
| 87. | How likely is it you will discuss the information at a formal meeting?                                                                                    | 1   2   3   4   5 |
| 88. | How important to you is discussing this information at a formal meeting? <b>Go to Q90</b>                                                                 | 1   2   3   4   5 |
| 89. | How important was discussing the information at a formal meeting?                                                                                         | 1   2   3   4   5 |
| 90. | <b>Have you been able to undertake further research and reading on the issues/recommendations raised in the evaluation information? If yes, go to Q93</b> | <b>yes    no</b>  |
| 91. | How likely is it you will undertake further research and reading on the issues/recommendations raised in the information?                                 | 1   2   3   4   5 |

## V. Use of Evaluation Information (continued)

92. How important to you is undertaking further research and reading on the issues/recommendations raised in the information? **Go to Q94** 1 2 3 4 5
93. How important was undertaking further research and reading on the issues/recommendations raised in the information? 1 2 3 4 5
94. **Have you been able to use the evaluation study as feedback on program implementation? If no, go to Q97** yes no
95. How likely is it you will use the evaluation study as feedback on program implementation? 1 2 3 4 5
96. How important is it to you to use the evaluation study as feedback on program implementation? **Go to Q98** 1 2 3 4 5
97. How important was it to you to use the evaluation study as feedback on program implementation? 1 2 3 4 5
98. **Have you been able to use the evaluation information to understand how the program could be improved? If yes, go to Q101** yes no
99. How likely is it you will use the evaluation information to understand how the program could be improved? 1 2 3 4 5
100. How important is it for you to use the evaluation information to understand how the program could be improved? **Go to Q102** 1 2 3 4 5
101. How important was it for you to use the evaluation information to understand how the program could be improved? 1 2 3 4 5
102. **Have you been able to use the evaluation study report to give the program credibility? If yes, go to Q105** yes no
103. How likely is it you will use the report to give the program credibility? 1 2 3 4 5
104. How important is it for you to use the report to give the program credibility? **Go to Q106** 1 2 3 4 5

## V. Use of Evaluation Information (continued)

- |      |                                                                                                                                         |                   |
|------|-----------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| 105. | How important was it for you to use the report to give the program credibility?                                                         | 1   2   3   4   5 |
| 106. | <b>Have you been able to use the information to establish a record of the program? If yes, go to Q109</b>                               | <b>yes   no</b>   |
| 107. | How likely is it you will use the information to establish a record of the program?                                                     | 1   2   3   4   5 |
| 108. | How important is it to you to use the information to establish a record of the program? <b>Go to Q110</b>                               | 1   2   3   4   5 |
| 109. | How important was it to you to use the information to establish a record of the program?                                                | 1   2   3   4   5 |
| 110. | <b>Have you been able to use the evaluation information to influence general policy decisions about the program? If yes, go to Q113</b> | <b>yes   no</b>   |
| 111. | How likely is it you will use the information to influence general policy decisions about the program?                                  | 1   2   3   4   5 |
| 112. | How important is it to you to use the information to influence general policy decisions about the program? <b>Go to Q114</b>            | 1   2   3   4   5 |
| 113. | How important was it to you to use the information to influence general policy decisions about the program?                             | 1   2   3   4   5 |
| 114. | <b>Has the evaluation information modified your views about the program? If yes, go to Q117</b>                                         | <b>yes   no</b>   |
| 115. | How likely is it that you will use the evaluation information to modify your views about the program?                                   | 1   2   3   4   5 |
| 116. | How important to you is the evaluation information in modifying your views about the program? <b>Go to Q118</b>                         | 1   2   3   4   5 |
| 117. | How important to you was the evaluation information in modifying your views about the program?                                          | 1   2   3   4   5 |
| 118. | <b>Have you been able to use the information to make major changes to the current program? If yes, go to Q121</b>                       | <b>yes   no</b>   |
| 119. | How likely is it you will use the information to make sweeping changes to the current program?                                          | 1   2   3   4   5 |

## V. Use of Evaluation Information (continued)

120. How important is it to you to use the information to make sweeping changes to the program? **Go to Q122** 1 2 3 4 5
121. How important to you was using the evaluation information to make major changes to the current program? 1 2 3 4 5
122. **Have you been able to use the evaluation study to justify the program to the community? If yes, go to 125** yes no
123. How likely is it you will use the evaluation study to justify the program to the community? 1 2 3 4 5
124. How important is it to you to use the evaluation study to justify the program to the community? **Go to Q126** 1 2 3 4 5
125. How important was it to you to use the evaluation study to justify the program to the community? 1 2 3 4 5
126. **Have you been able to use the evaluation information to legitimate and justify what has been done in the program? If yes, go to Q129** yes no
127. How likely is it you will use the evaluation information to legitimate and justify what has been done in the program? 1 2 3 4 5
128. How important is it to you to use the evaluation information to legitimate and justify what has been done in the program? **Go to Q130** 1 2 3 4 5
129. How important was it to you to use the evaluation information to legitimate and justify what has been done in the program? 1 2 3 4 5

**V. Use of Evaluation Information (continued)**

**130.** Could you describe any other ways in which you have or might use the evaluation information?

**131.** How important to you is this type of use for the program? 1 2 3 4 5

**132.** Do you have any other comments you'd like to make before we close?

**Thank-you for your time and for your responses**



## **APPENDIX C**

1. Information Letter for Interviewees
2. Consent Form to be signed by Interviewee
3. Evaluator Interview Schedule I
4. Evaluator Interview Schedule II

Dear \_\_\_\_\_

I look forward to including your responses to my interview questions in my study data.

As you know, I am a Ph.D. student at Murdoch University researching the use of evaluations and their information. This current study extends research completed as part of my Master's Degree considering the influences on the use of evaluation information such as the nature of the host organisation, the stakeholders and the evaluation process.

All involvement I have with any program, its staff and evaluation will be conducted in accordance with the ethical guidelines for practice advocated by the Australasian Evaluation Society, and the Murdoch University research ethics code of practice regarding human research. I will undertake precautions to ensure no names or other information that may identify you are reported in any informal paper or publication linked to this study. All completed interview schedules will be coded and stored in a secure area.

I will make a summary of the study findings available to you when I have finished. Clearly, sound research considering the evaluation process and the factors influencing that process should provide insightful information useful to the organisation funding the evaluation and the program evaluated.

Should you think of any questions you would like to ask me regarding this research at a later time my contact details are below.

Kind Regards,

Sheri Hudson-Mabbs  
**MURDOCH UNIVERSITY**  
*24 Aberfoyle Heights*  
*Currambine WA 6028*  
*(08) 9305 3847*  
[bigskies@iprimus.com.au](mailto:bigskies@iprimus.com.au)

## RESEARCH CONSENT FORM

### **Project Title: Factors Influencing the Use of Evaluation Information: A Longitudinal Study**

I am a Ph.D. student at Murdoch University investigating the influence of factors on the use of program evaluation information in organisations. The purpose of the study is to identify more efficient and useful evaluation practices and processes.

You can help in this study by consenting to be interviewed. Contained in the interview schedule are questions that relate to the evaluation of the XXXXXX program, and the use of its information. You may decide to withdraw your consent at any time.

All information given during the interview will be confidential. No names or other information that may identify you will be used in any publication arising from the study. Feedback on the study will be provided to you.

If you are willing to participate in this study, could you please complete the details below? If you have any questions about this project please feel free to contact either myself, Sheri Hudson-Mabbs, on 9305 3847, Associate Professor Ralph Straton on 9360 6995, or Dr. Richard Cummings on 9360 2354.

My supervisors and I are happy to discuss with you any concerns you may have on how this study is conducted, or you can contact Murdoch University's Human Research Ethics Committee on 9360 6677.

\*\*\*\*\*

I (the participant) have read the information above. Any questions I have asked have been answered to my satisfaction. I agree to take part in this activity; however I may change my mind and withdraw at any time.

I understand that all information provided will be treated as confidential and will not be released by the investigator unless required to do so by law.

I agree for this interview to be taped.

I agree that research data gathered for this study may be published provided my name or other information which might identify me is not used.

Participant/Authorized Representative's Name:	Investigator's Name:
Participant/ Authorized Representative's Signature:	Investigator's Signature:
Date:	Date:

## **EVALUATOR INTERVIEW SCHEDULE I**

**SHERI HUDSON-MABBS  
MURDOCH UNIVERSITY  
OCTOBER 1, 2009**

### **INTRODUCTION**

**THE PURPOSE OF THIS INTERVIEW IS TO GATHER INFORMATION FROM YOU ABOUT YOUR EVALUATION AND RESEARCH BACKGROUND AND YOUR PRACTICE AS AN EVALUATOR. IN ADDITION, SOME OF THE INTERVIEW QUESTIONS FOCUS ON YOUR ROLE IN THE EVALUATION TEAM OF THE \_\_\_\_\_, HOW YOU BECAME INVOLVED, AND YOUR EXPERIENCE OF THE EVALUATION.**

<b>Name:</b>	
<b>Date:</b>	
<b>Title:</b>	
<b>Program:</b>	
<b>Organisation:</b>	

## SECTION A

### Training, Experience, Commitment, and Reflections on the Evaluation Study

First, I am going to ask you some questions about your role in the evaluation of the \_\_\_\_\_ program.

1. What role do you play in the evaluation team (e.g. researcher, advisor, team leader, etc...)?

---

---

---

---

- 2.a. How is it that you became involved in the evaluation of this particular program (e.g. employee, tenderer, specialist advisor, etc...)?

---

---

---

---

- b. If your involvement with the evaluation of this program was the result of a selection process, do you feel there was any reason in particular that you were selected (e.g. relevant work history or experience, pertinent training, etc...)?

---

---

---

---

## SECTION B

**The next few questions focus on your evaluation/research background in terms of training, education, work experience and commitment to the field.**

- 3a. What education and training have you undertaken that has influenced your thinking and practice regarding evaluation and research (e.g. relevant workshops and courses attended, higher education, postgraduate study, etc...)?

**INTERVIEWER:** *PROMPT QUESTION WITH COLUMN HEADINGS OF TABLE.*

<b>Workshop, course, unit, university degree, etc..., title and duration.</b>	<b>Briefly summarise how these educational and training courses influenced your thinking and practice regarding research and evaluation.</b>	<b>Approximately what length of course, workshop, etc..., time focused on research/evaluation theory and skills?</b>

- 3b. What conferences and seminars have you attended in the past five years hosted by associations or societies concerned with evaluation or research practice, such as the Australasian Evaluation Society.?

**INTERVIEWER:** *PROMPT QUESTION WITH COLUMN HEADINGS OF TABLE .*

<b>Year of conference attendance?</b>	<b>Association or Society hosting conference?</b>	<b>Did you present any papers?</b>	<b>Did you play a role in organising or running the conference? If so, what did you do?</b>



4. Have you conducted any research studies in the past five years? If so, can you tell about one which you feel has influenced your thinking or practice regarding social research the most?

**INTERVIEWER:** *USE THE QUESTIONS IN THE TABLE TO PROMPT RESPONSE INFORMATION. COPIES OF THE TABLE ARE PRINTED ON THE PINK SHEETS AT THE BACK OF THE QUESTIONNAIRE FOR ADDITIONAL STUDIES.*

<b>a. What was the purpose of the study?</b>
<b>b. How was the information to be collected prioritised? What was the study to achieve?</b>
<b>c. What methods were used to gather the information?</b>
<b>d. In what way were you involved with the study? What was your primary role?</b>
<b>e. How long was the study?</b>
<b>f. How long were you involved with the study?</b>

5. Can you tell me about one evaluation study have you been involved with recently?

**INTERVIEWER:** *USE THE QUESTIONS IN THE TABLE TO PROMPT RESPONSE INFORMATION. COPIES OF THE TABLE ARE PRINTED ON THE BLUE SHEETS AT THE BACK OF THE QUESTIONNAIRE FOR ADDITIONAL STUDIES*

**a. What program was the focus of the evaluation?**

**b. What was the purpose of the study?**

**c. How was the information to be collected prioritised?  
What was the study to achieve?**

**d. What methods were used to gather the information?**

*Item 5 continued on next page.....*

*Item 5 continued from previous page....*

**e. In what way were you involved with the study?  
What was your primary role?**

**f. How long was the study?**

**g. How long were you involved with the study?**

**h. The logic or theory of the program is the underlying assumptions about how a program will work to achieve intended outcomes. Did the evaluation study consider the program's logic or theory in any way?  
If yes, how so (e.g. theory explication, theory use in decisions regarding study form, approach or methods, theory use in information analysis).**

6. Now, I would like to ask you a few questions about the evaluation of the \_\_\_\_\_ program.

**a. What is the purpose of the study?**

**b. How was the information to be collected prioritised?  
What is the study to achieve?**

**c. What methods are being used to gather the information?**

**d. In what way are you involved with the study?  
What is your primary role?**

*Item 6 continued on next page.....*

*Item 6 continued from previous page....*

**e. How long is the study expected to be?**

**f. How long do you expect to be involved with the study?**

**g. The logic or theory of the program is the underlying assumptions about how a program will work to achieve intended outcomes.  
Will the evaluation study consider the program's logic or theory in any way?  
If yes, how so (e.g. theory explication, theory use in decisions regarding study form, approach or methods, theory use in information analysis).**

## SECTION C

### Evaluator Role

7. The purpose of this section is to gain insight into the relative importance you place on various roles an evaluator plays in the course of an evaluation. To do this allocate 20 points across the roles below in accordance with your thinking regarding which are the most important roles of an evaluator (e.g. 10 points to one, 5 points to another, 2 points to another, etc...). Award the most important roles the higher number of points. No two roles should have been awarded the same number of points. All roles must be allocated some points. Once you have distributed your points between the roles please check that they total 20.

The following are roles adopted by the evaluator.

**A judge of quality** independently determining the merit or worth of the program. \_\_\_\_\_

**A researcher** focused on the evaluation task and concerned With the technical issues involved with the study. \_\_\_\_\_

**An educator** intent on encouraging those involved with the program evaluation to learn from the evaluation experience. \_\_\_\_\_

**A facilitator** guiding those involved through the evaluation process.. \_\_\_\_\_

**A learning partner** intent on sharing the evaluation process As a learning journey with others involved in the evaluation.. \_\_\_\_\_

## SECTION D

### Evaluator Role

The purpose of this section is to seek your views on of the various roles listed that an Evaluator might take on in the course of an evaluation. Scores for the importance of two evaluation practices are required for each role. These practices are *Ongoing Communication with Stakeholders* and *Evaluator Control of the Evaluation Process*. For instance, with regards to a 'Judge of Quality', how important is it that the evaluator establish *Ongoing Communication with Stakeholders*? Please respond to each item on the scale to the left.

	Not at All				A Great Deal
In the evaluator role of:	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
<b>8.a. A judge of quality</b> independently determining the merit or worth of the program how important is,					
<i>Ongoing Communication with Stakeholders?</i>	1	2	3	4	5
<i>Evaluator Control of the Evaluation Process?</i>	1	2	3	4	5
<b>b. A researcher</b> focused on the evaluation task and concerned with the technical issues of the study how important is,					
<i>Ongoing Communication with Stakeholders?</i>	1	2	3	4	5
<i>Evaluator Control of the Evaluation Process?</i>	1	2	3	4	5
<b>c. An educator</b> intent on encouraging those involved to learn from the evaluation experience how important is,					
<i>Ongoing Communication with Stakeholders?</i>	1	2	3	4	5
<i>Evaluator Control of the Evaluation Process?</i>	1	2	3	4	5
<b>d. A facilitator</b> guiding those involved through the evaluation process how important is,					
<i>Ongoing Communication with Stakeholders?</i>	1	2	3	4	5
<i>Evaluator Control of the Evaluation Process?</i>	1	2	3	4	5
<b>e. A learning partner</b> intent on sharing the evaluation process as a learning journey with others involved in the evaluation how important is,					
<i>Ongoing Communication with Stakeholders?</i>	1	2	3	4	5
<i>Evaluator Control of the Evaluation Process?</i>	1	2	3	4	5

## SECTION E

### Evaluator Practice

**INTERVIEWER:** The questions in this section have been included to gather information from you regarding your thinking about evaluator practice. On the scale provided please indicate the extent to which you agree with the items below.

	Not at All	A Great Deal			
In the evaluator role an evaluator should:	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
9. Take account of stakeholders views of the program (e.g. able to put themselves in the stakeholder's shoes)?	1	2	3	4	5
10. Work to establish communication pathways with stakeholders?	1	2	3	4	5
11. Be open to negative feedback about the evaluation process?	1	2	3	4	5
12. Be flexible in undertaking an evaluation?	1	2	3	4	5
13. Be aware of the history influencing the program being evaluated?	1	2	3	4	5
14. Be aware of the influence of the environment within which the program is set (e.g. relevant policy, availability of resources, level of stakeholder authority...)?	1	2	3	4	5
15. Be aware of the influence of program providers/practitioners on the program being evaluated?	1	2	3	4	5
16. Be aware of the influence of program clients/recipients on the program being evaluated?	1	2	3	4	5
17. Understand the rationale, logic or theory of the program? <i>Remember, the logic, theory or rationale of the program is the underlying assumptions about how a program will work to achieve intended outcomes.</i>	1	2	3	4	5
17. Take into account the rationale, logic or theory of the program when undertaking the evaluation?	1	2	3	4	5
19. Encourage stakeholders to explore the rationale, logic or theory guiding the program?	1	2	3	4	5
1. Take into account the stakeholders' thinking regarding the rationale, logic, or theory guiding the program?	1	2	3	4	5



21. a. When planning the evaluation an evaluator should  
(Please select only one):

Integrate into the evaluation the ideas stakeholders have  
regarding the evaluation.

☐

or

Follow their ideas and strategies for the evaluation process  
in preference to the ideas stakeholders have regarding the  
evaluation?

☐

**You've indicated it is more important for an  
evaluator to ....(repeat above selection).**

- b. How important do you think it is to do this?                      1    2    3    4    5

**INTERVIEWER:    The    following    questions    are    about    stakeholder  
involvement.**

Not at All                      A Great Deal

**To what extent, in your opinion, should:                      1    2    3    4    5**

22. Stakeholders be involved in each of the following  
areas:

- a. defining what the study focuses on?                      1    2    3    4    5

- b. exploring the logic/theory or rationale underpinning  
the program?                      1    2    3    4    5

(Underlying assumptions about how a program  
will work to achieve intended outcomes.)

- c. developing the objectives of the study?                      1    2    3    4    5

- d. identifying what information is to be collected?                      1    2    3    4    5

- e. deciding how the information is to be collected?                      1    2    3    4    5

- f. identifying who the information will be collected from?                      1    2    3    4    5

- g. determining how the information will be reported?                      1    2    3    4    5

23. Are there any further comments you would like to make?

***Thank-you for your time***

*ADDITIONAL COPIES OF THE TABLE TO RECORD RESPONSES REGARDING INVOLVEMENT IN EVALUATION STUDIES.*

<p><b>* What was the purpose of the study?</b></p>
<p><b>* How was the information to be collected prioritised? What was the study to achieve?</b></p>
<p><b>* What methods were used to gather the information?</b></p>
<p><b>* In what way were you involved with the study? What was your primary role?</b></p>
<p><b>* How long was the study?</b></p>
<p><b>* How long were you involved with the study?</b></p>
<p><b>* Did the study consider the logic or theory of the program in any way? If yes, how so (e.g. theory explication, theory use in decisions regarding study form, approach or methods, theory use in information analysis).</b></p>

*ADDITIONAL COPIES OF THE TABLE TO RECORD RESPONSES REGARDING INVOLVEMENT IN EVALUATION STUDIES.*

<p><b>* What was the purpose of the study?</b></p>
<p><b>* How was the information to be collected prioritised?</b>  <b>What was the study to achieve?</b></p>
<p><b>* What methods were used to gather the information?</b></p>
<p><b>* In what way were you involved with the study?</b>  <b>What was your primary role?</b></p>
<p><b>* How long was the study?</b></p>
<p><b>* How long were you involved with the study?</b></p>
<p><b>* Did the study consider the logic or theory of the program in any way?</b>  <b>If yes, how so (e.g. theory explication, theory use in decisions regarding study form, approach or methods, theory use in information analysis).</b></p>

**CODE:**\_\_\_\_\_

**Evaluator Interview Schedule II**

**Sheri Hudson-Mabbs  
Murdoch University  
October 1, 2009**

<b>Name:</b>	
<b>Date:</b>	
<b>Title:</b>	
<b>Program:</b>	
<b>Organisation:</b>	

## **Introduction**

**The purpose of this interview schedule is to collect information from people, such as yourself, responsible for undertaking the evaluation of the \_\_\_\_\_ Program. In particular, questions included in this interview focus on your perception of the evaluation process and the information that has resulted from it.**

## **Evaluator Characteristics Interview Schedule II**

First, I would like to ask you some questions about your experience of this evaluation study. The purpose of the questions below is to gather information about your perception of the evaluation process.

1. Thinking about the evaluation of the \_\_\_\_\_ program,

a. In your opinion, what was the purpose of the study?

b. How was the information to be collected prioritised?

c. What was the study to achieve?

d. What methods were used to gather the information?

e. In what way were you involved with the study?

f. What was your primary role?

g. How long was the study?

h. How long were you involved with the study?

i. The logic or theory of the program is the underlying assumptions about how a program will work to achieve intended outcomes. Did the evaluation study consider the program's logic or theory in any way?

If yes, how so (e.g. theory explication, theory use in decisions regarding study form, approach or methods, theory use in information analysis).

The next few questions ask you for a response on a five point scale. I will direct you to the appropriate scale for each question on the page I have given you. After you have scored a response for each question I will ask you why you feel as you do with regard to each question.

	<u>Not Appropriate</u>			<u>Very Appropriate.</u>	
2 a. To what extent do you feel your skills are appropriate to this evaluation?	1	2	3	4	5

b. Why or why not?

-----

-----

-----

-----

-----

-----

-----

	<u>Not Competent</u>			<u>Very Competent</u>	
3.a. To what extent do you feel the evaluation team was competent to undertake the evaluation study?	1	2	3	4	5

b. Why or why not?

-----

-----

-----

-----

-----

-----

-----



		<u>No Latitude</u>			<u>A Great Deal of Latitude</u>	
		1	2	3	4	5
4.a.	To what extent do you feel the evaluation team was given sufficient latitude to exercise sound professional judgment in undertaking the evaluation?					

b. Why or why not?

-----

-----

-----

-----

-----

-----

-----

-----

		<u>Very Negative Experience</u>			<u>Very Positive Experience</u>	
		1	2	3	4	5
5.a.	Thinking about the evaluation of the _____ program, to what extent do you consider the evaluation to have been a negative or positive professional experience?					

b. Why do you feel this way?

-----

-----

-----

-----

-----

-----

-----

-----

6. In what ways do you think primary stakeholders involved in the evaluation

were influenced by the evaluation process?

---

---

---

---

---

---

7. In what ways do you think primary stakeholders involved in the evaluation were influenced by the evaluation findings?

---

---

---

---

---

---

8. In what ways do you feel the evaluation information has or will be put to use?

---

---

---

---

---

---

9. Do you have any final comments you'd like to make?

-----

-----

-----

-----

-----

-----

**Thank-you for your time and for your responses**

## SCALES

Not Appropriate      Very Appropriate

1    2    3    4    5

Not Competent      Very Competent

1    2    3    4    5

No Latitude      A Great Deal  
of Latitude

1    2    3    4    5

Very Negative  
Experience      Very Positive  
Experience

1    2    3    4    5

## **APPENDIX D**

1. Instrument for developing a scale to rate the degree of program stake or interest of various stakeholder roles.

Hello

I am developing a scale to rate the degree of program stake or interest of various stakeholder roles. Please rank your perception of the degree of program interest of each stakeholder role to the right. No two roles can be given the same rank. All 10 roles must be ranked from 0 to 10. Your assistance with this task is appreciated.

	<b>STAKE IN PROGRAM</b>	<b>YOUR RATING 0 - 10</b>
1	No immediate stake in program	
2	Supervise/oversee program- Department/ Organisation Level	
3	Manage/co-ordinate program- Department/ Organisation Level	
4	Oversee/supervise program- Local Delivery	
5	Manage/co-ordinate program- Local Delivery	
6	Support/administrative- Program	
7	Delivery- Program	
8	Participant/Client	
9	Program Sponsor	
10	Member of Executive responsible for program.	

Recoded values for Program Stake to include “0” value.

<b>CODE</b>	<b>ROLE DESCRIPTION</b>
0	No immediate stake in program
1	Supervise/Oversee program (Department Level)
2	Manage/ Co-ordinate Program (Department Level)
3	Oversee/ Supervise Program- (Local Level)
4	Manage/ Co-ordinate Program- (Local Level)
5	Support/ Administrative – Program
6	Delivery- Program
7	Participant/Client
8	Program Sponsor
9	Member of Executive responsible for program